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

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

March 3, 2026

Abstract

Using thousands of essays written by 11-year-olds in 1969, we construct an index measuring children's conformity to gender norms. We link this index to outcomes over the life-cycle. Conditional on age-11 covariates, a one standard deviation increase in our index predicts a 4% decline in lifetime earnings for girls, associated with lower wages and fewer hours worked. We find no statistically significant association for boys. Education, occupation, and family formation account for 40% of the earnings decline observed among girls.

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

[¶]Thanks to Barbara Biasi, Enda Hargaden, Stelios Michalopoulos, and seminar participants at Bonn, Brown, Cambridge, LSE, Manchester, IFAU, University College Dublin, the IFS Conference Intergenerational Inequality, the NBER Children Program Spring Meeting, and the Zurich Text-As-Data conference for comments. We are grateful to Richard Gong, Peyton Crevasse, Dylan Miner, and Alexander-Minh Vu 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, Josh Leffler, Ethan Liu, Nastaran Moghimi, Eleri Phillips, Charlotte Townley, and Teddy Tawil for correcting spelling mistakes in the essays. Funding from 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. This paper uses observational data only and is not accompanied by a pre-analysis plan. Opinions, conclusions, and errors are solely those of the authors.

1 Introduction

Despite a narrowing gender earnings gap, considerable differences in earnings between men and women remain (Goldin, 2021). Women's conformity to prevailing gender norms can potentially play a substantive role in explaining such earnings differences (Fortin, 2005; Bertrand, 2020; Boelmann et al., 2025).[1] In this paper, we use novel text data — thousands of essays written by 11-year-old children in 1969 – to construct an index of gender conformity, which measures the extent to which children imagine their future to conform to gender norms at the time. We link these essays to panel data that tracks their writers' lives from birth to retirement.

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 feminine conformity in their essays have 4% lower lifetime earnings due to both lower wages and fewer hours worked. In contrast, there is no significant association between our measure and lifetime earnings for boys. Text that relates to hobbies and home life is particularly predictive of lifetime earnings for girls. Overall, 40% of the association between earnings and conformity for girls can be explained by a combination of lower educational attainment, selection into lower-paid occupations, and family formation of girls who display stronger feminine conformity.

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. While early analyses described a sub-sample of the essays (Richardson et al. (1976), Elliott (2010)), the full sample of essays has 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

[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. Relatedly, feminine conformity refers to adhering to prevailing female gender norms. 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.

(Pongiglione et al., 2020; Wielgoszewska et al., 2022; Kern et al., 2022).

In this paper, we extract a measure of gender conformity from children's essays using natural language processing. The assumption underlying our measurement of conformity is that girls (boys) conform more strongly to prevalent gender norms if the content and style of their essay are associated with females (males). For example, describing family and housework (as opposed to career and football) is more typically feminine. Whilst we account extensively for measurable differences in circumstances which may constrain girls' opportunities (e.g. due to family background or geographic differences), we remain agnostic as to whether the association between gender conformity and earnings arises from differential preferences, including those shaped by social identity (Akerlof and Kranton, 2000), or from the differential constraints that the review by Olivetti et al. (2024) emphasizes. As they discuss, women and men typically face differential trade-offs between family and career rooted in gender roles within the household that are shaped by wider societal norms, with implications for job sorting, job search, and earnings. We discuss how both preferences and constraints could rationalize our findings using a model of education choice and time allocation.

To estimate a measure of feminine conformity, we first manually correct over 100,000 spelling mistakes in the essay data. Then, to measure how 'feminine' or 'masculine' individual words were at the time the essays were written, we apply a method proposed by Kozlowski et al. (2019) and train a Word-Embedding Model using the Google Books Ngram Corpus (Lin et al. (2012)). We use over 140,000 books written in British English during the first 21 years of the writers' lives (between 1958 and 1978). The Word-Embedding Model infers the femininity of words from their association with unambiguously gendered words like "women", "girl", "she", "her" etc. Using this measure of how feminine individual words are, we measure gender conformity of an essay using the prevalence of these gendered words in the children's essays.

Holding age-11 measured skills and other characteristics constant, girls who write essays with one standard deviation stronger feminine conformity have 4% lower lifetime earnings. This is due to both lower wages and fewer hours of work across the lifecycle. There is no statistically significant association for boys. We do not interpret these associations as causal, but instead as conditional associations that are robust to the inclusion of a large number of additional covariates.

To uncover what explains the association between gender conformity and lower lifetime earnings for girls, we perform mediation analysis. This allows us to estimate how much of the association between lifetime earnings and feminine conformity is explained by feminine 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 1.2 percentage point decrease in the probability of entering a professional occupation. Girls who conform more strongly to feminine gender norms are also more likely to have children. 40% of the association between lifetime earnings and gender conformity is mediated by education, occupation, and family formation.

In the final part of the paper, we consider what determines gender conformity. We find that girls with better cognitive 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 children internalize prevailing gender norms. We measure this using information on whether girls 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 (mostly) feminine gender norms and outcomes. The first is to proxy a woman's feminine adherence to norms using those prevalent in her country of birth, since these norms have been found to persist across generations (Alesina et al., 2013). Examples of this approach include Fernández and Fogli (2009), Boelmann et al. (2025), and Ichino et al. (2026), who use data on migrants to the US, within Germany, and to Sweden respectively, to study how gender norms from a woman's country of origin influence labor market and household decisions. These papers find that such norms have significant explanatory power for women's labor supply and the division of household labor, even after migration to countries 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 measures 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 household income exhibits a sharp drop when it exceeds 0.5 among 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. Bursztyn et al. (2017) show that gender norms pose a trade-off for single women too. Using two field-experiments among MBA students in the US, they show that single women avoid career-enhancing actions which signal traits such as ambition and assertiveness, so as to increase their chance of finding a male partner. 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 index, constructed using open-ended text in childhood essays, is distinct and has three advantages relative to the above approaches. First, we measure gender conformity at a young age, *before* career and family decisions are made. Thus, our study is less susceptible to reverse causality resulting from labor market outcomes affecting reported gender attitudes or justification bias (Black et al., 2017). Second, because respondents were asked to write about their future and not prompted to consider 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)), which can affect studies of gender norms based on survey measures. Third, we link our measure of gender conformity to skills and circumstances in early childhood as well as outcomes through different stages of adulthood. This allows us to study both the determinants and consequences of gender conformity.

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 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 using published opinions from circuit courts in the US to obtain measures 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. 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 introduces our data. Section 3 discusses how we estimate gender conformity. Section 4 contains our results. Section 5 concludes.

2 Data

Our data come from the National Child Development Study (NCDS).[2] The data cover almost all children born in Great Britain in the third week of March 1958 and follows them to this day. The purpose of the NCDS survey was to gain a broad understanding of children's health and development. 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 throughout adulthood.

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

The initial survey at birth was followed by subsequent surveys at ages: 7, 11, 16, 23, 33, 42, 46, 50, and 55.[3] The data from childhood includes measures of cognitive and non-cognitive skills, family circumstances, and parental income. Later waves of the study record educational attainment, family formation (child-bearing, marriage etc.), earnings, and hours of work.

2.1 The Essays

The age 11 (1969) NCDS survey included a set of written exercises, conducted in cohort members' school classrooms. It included assessments of reading, arithmetic, and general ability, as well as questions about their interests. In addition, NCDS cohort members were assigned 30 minutes to write an essay on what they imagined their lives would be like at 25 (Centre for Longitudinal Studies, 2018). The essay 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.*

The essay was administered at the end of a test booklet that included the cognitive assessments, with a total allocated time for the booklet to be 3 hours. Testing was supervised by class teachers, and it was recommended that it be spread over two days (Richardson et al., 1976).

Panel A of Table 1 provides summary statistics on the full sample of 10,511 essays, which we then split into the 5,091 essays written by girls and 5,420 written by boys. Panel B presents summary statistics for our analysis sample of 4,056 girls and 4,030 boys.

Relative to boys, girls write essays that are about 25% longer on average, and that have fewer spelling mistakes. The length of essays and number of spelling mistakes is similar in our analysis sample as in the full sample of essays.

Figure 1 shows the 80 most frequently used words by gender (excluding articles and pronouns). The essays span diverse topics including sport, domestic chores, career choice, and family life, with clear gender differences in emphasis.

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

Table 1: Describing the Essays

	Mean	Standard Deviation	25th Percentile	50th Percentile	75th Percentile
Panel A: Full Essay Sample					
Full Sample					.
Essay Length (words)	198.7	107.1	123.0	182.0	257.0
Spelling Mistakes	10.0	8.5	4.0	8.0	13.0
Girls Full Sample					.
Essay Length (words)	223.7	113.8	143.0	207.0	286.0
Spelling Mistakes	9.4	7.9	4.0	7.0	12.0
Boys Full Sample					.
Essay Length (words)	175.3	94.6	108.0	160.0	226.0
Spelling Mistakes	10.7	8.9	5.0	8.0	14.0
Panel B: Analysis Sample					
Full Analysis Sample					.
Essay Length (words)	202.1	107.4	126.0	186.0	260.0
Spelling Mistakes	10.0	8.4	4.0	8.0	13.0
Girls Analysis Sample					.
Essay Length (words)	225.7	114.2	144.0	211.0	288.0
Spelling Mistakes	9.3	7.9	4.0	7.0	12.0
Boys Analysis Sample					.
Essay Length (words)	178.3	94.3	112.0	163.0	228.0
Spelling Mistakes	10.6	8.8	5.0	8.0	14.0

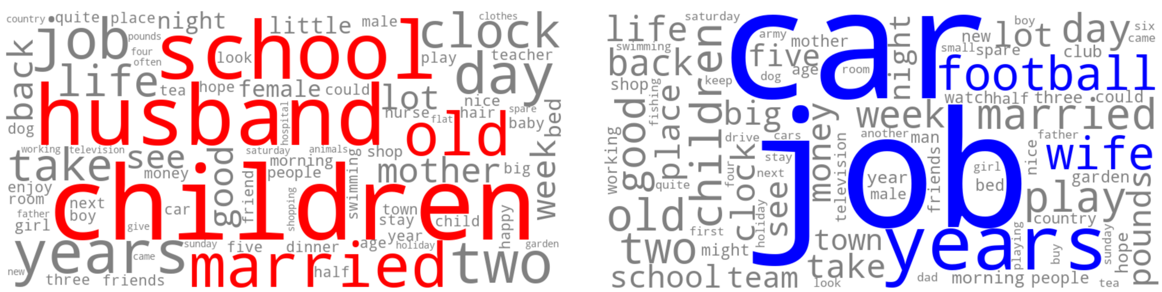
Notes: The full sample is 10,511 essays, which is the total number of digitized essays in the NCDS data. There are 5,091 essays written by girls and 5,420 written by boys. Our analysis sample of essays refers to the 4,056 (4,030) essays written by girls (boys), for whom we have data in adulthood and at least 1 month of employment information. We report essay length (after spelling correction) and spelling mistakes. The difference in mean essay length between those in our analysis sample versus those excluded is statistically significant at the 1% level for boys and at the 5% level for girls (p-value is 0.000 for boys and 0.006 for girls) – essays in our analysis sample are longer, on average. The p-value of the difference in mean spelling mistakes for those included versus excluded from our analysis sample is 0.786 for boys and 0.831 for girls.

2.2 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, geographic variables, 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,

Figure 1: Eighty Most Frequent Words, by Sex of Writer



(a) 80 Most-used (by females) Words

(b) 80 Most-used (by males) words

Notes: The figure shows the 80 most frequently used male and female words in our analysis sample (8,086 essays with ≈ 17,000 distinct words). The blue highlighted words are the five most-used male words, and the highlighted red words are the most-used female words. Words which occur more often have a larger font size.

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., 2024; Todd and Zhang, 2020). Non-cognitive skills capture a child's social adjustment and temperament. We have several teacher-reported measures of cohort members' behaviors at age 11, which we consider noisy measures of latent non-cognitive skills. We use the same measures of internalizing and externalizing behavior as Papageorge et al. (2024). To form indexes from these measures, we use the same latent variables approach as we do for cognitive skills to construct the indices for internalizing and externalizing behavior. Details on this procedure are also in Appendix B.2.[4]

Geographic Variables The survey records region of residence (of which there are 11) at age 11. We use these data to estimate region-level female employment and university-attendance

[4]More specifically, for externalizing behavior, we use 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 socializing skill correlate positively. For internalizing behavior, we use measures which record whether a child is anxious for acceptance or is hostile towards adults or children, writes off adults and acts without regard for others.

rates using information on the mothers of NCDS cohort members using a leave-one-out measure for each cohort member, excluding the mother of the cohort member of interest. The survey also contains more detailed county of residence: 'county' (of which there are 111) at age 16. 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, although all our results are robust to using age 11 region of residence.

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, the age 11 survey includes 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.

2.3 Adulthood Data

At ages 23, 33, 42, 50, and 55 we observe outcomes in adulthood.

Earnings We observe gross weekly earnings in every survey wave, which we convert to a monthly measure. We impute missing earnings by regressing earnings on an individual fixed effect and educational attainment interacted with a survey wave fixed-effect. Using the NCDS activity histories (which record employment status and type of work (part- or full-time) for each month between ages 23 and age 55 surveys), we impute monthly earnings for the period between interviews, setting 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. For individuals who never worked between ages 23-55, we set lifetime earnings to the first percentile of earnings for each gender (which is non-zero), allowing those individuals to remain in the sample when we take logarithms.

Wages We define a measure of hourly wages at ages 23, 33, 42, 50, and 55, by dividing weekly earnings by weekly hours of work. Using the same fixed-effects procedure as for earnings, we impute hourly wages for non-interview months across all 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 We compute the share of time spent in employment using the NCDS activities history by finding the fraction of months cohort members spent either in full- or part-time employment. Total hours sums over all work hours between 23 and 55, computed assuming 40 hours/week for full-time work, 20 hours/week for part-time work, and 0 for those not working.

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.[5]

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).

[5] O-levels were exams typically taken at age 16; A-levels are typically taken at age 18.

Quality of Life and Health At age 55, cohort members were asked about their quality of life, mental health and physical health. The NCDS data includes a derived index for each of these three outcomes. Details on their construction are given in Appendix B.5. For our analysis, we standardize these indexes to have mean zero and variance one.

2.4 Sample Selection

The NCDS panel data contains 10,511 essays. For the purpose of estimating our index of gender conformity, we use all available essays.[6] Our analysis sample, which we use across all our regressions, is formed of individuals: (i) who wrote essays, (ii) who were surveyed at least once in adulthood, and (iii) whose employment status was observed at least once in adulthood. Our final sample consists of 8,086 individuals, of whom 4,056 are girls and 4,030 are boys. Appendix B.1 provides details, including a breakdown of attrition in the NCDS surveys.

3 An Index of Feminine Conformity

A key contribution of this paper is our novel measure of feminine conformity, constructed from the age-11 NCDS essays. We measure feminine conformity based on the extent to which words in each essay are associated with femininity. Our word embedding model identifies a word as feminine if it frequently appears in context with unambiguously gendered words like "she," "her," and "woman," in a large corpus of books. For example, words like 'cooking', 'baby', or 'sister' would be considered more feminine than writing about work and football, because, at least, in books written around the time our respondents wrote their essays, the former words more commonly appeared alongside "she," "her," and "woman".

This section describes construction of our index of feminine conformity, provides descriptive information about our index, and provides evidence of its validity. The following two points are noteworthy for interpreting what we measure. First, we capture the extent to which children's imagined futures conform to gender norms prevalent *at the time the essays were written* (in

[6] Goodman et al. (2017) show that 13,675 of the 15,335 NCDS 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. Ultimately, among 13,675 essays, 10,551 were successfully digitized; the remainder could not be digitized due to the microfiche being missing or illegible.

1969). Because the roles, norms, and activities that are typically associated with being feminine differ across time (Kleven and Landais, 2017; Kozlowski et al., 2019), we evaluate the gender associations of an essay's content relative to texts written in Britain between 1958 and 1978.[7] Second, our feminine conformity index captures how much children's attitudes towards their *own* future are in line with traditional norms; it does *not* capture what they think other men or women should do in general.[8]

3.1 A Roadmap for Estimating Gender Conformity

Our procedure has six steps, which we list here and explain in the subsequent subsections. First, we correct over 100,000 spelling errors across the 10,511 essays. Second, we represent each essay as a set of word counts. Third, using a Word Embedding Model, we assign each word in the essays a vector value that encodes its meaning and contextual associations, so that similar words, and words used in similar contexts, have similar vectors. Fourth, using pairs of unambiguously gendered words (e.g., woman–man, girl–boy, female–male) and the Word Embedding Model, we construct a "gender dimension". This is the average of 10 vectors, each representing ('feminine word'–'masculine word') and essentially gives us an axis in the vector space that captures the direction from masculine to feminine. Fifth, we project the vector for each word in the essays onto the gender dimension to determine its femininity or masculinity. Sixth, we multiply projection values with how many times each word appears in an essay and sum over words. We then standardize this to mean zero and variance one, yielding our index of gender conformity. In what follows, we explain this procedure in more detail.[9]

[7] Another recent 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.

[8] This contrasts with recent work on eliciting individuals' perceptions of social norms (Gauri et al., 2019; Bursztyn et al., 2020; Dhar et al., 2022).

[9] Appendix C provides justification and validation on each of these steps, with specifics on implementation and how to replicate our procedure.

3.2 Estimating Gender Conformity

3.2.1 Step 1: Digitization, spelling correction, and pre-processing

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

To make our essays usable for text analysis, we correct spelling mistakes in the essays. First, we identify (potentially) 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 indeed misspelled. Correcting misspelled words individually allows us to maintain the essay's original structure and message which spell-check software may misinterpret.

3.2.2 Step 2: Representation of essays as counts of words

After removing punctuation and stop words (e.g., articles and pronouns), we count individual words (1-grams). To capture negation, we also include 2-grams beginning with "not" or "no," allowing us to distinguish, for example, "married" or "wife" from "not married" or "no wife". We discuss how we deal with negations in Step 3.

Table 2: Example - Bag of Words

friend	**husband**	**married**	**not married**
2	1	1	1

Table 2 illustrates our Bag of Words procedure using the sentences: "My friend is not married. Does your friend have a husband?" We drop stop words ("my, is, does, your, have, a"), count 1-grams ("friend, married, husband"), and add the 2-gram "not married." Other 2-grams (e.g., "friend married") are ignored. Our representation of the essays follows the same logic. We construct a matrix with over 18,000 1-grams (and 2-gram negations) and associated counts per essay.

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

We train a Word-Embedding Model (WEM) to assign a numerical vector to each 1-gram in our Bag of Words matrix. WEMs represent words as high-dimensional, real-valued vectors. Words with similar meanings and/or association receive similar vector values, allowing relationships between them to be quantified using linear algebra (Mikolov et al., 2013).

Several off-the-shelf WEMs are available. However, we are not aware of any that are specific to the time period when the essays were written. Since word meanings and associations potentially change over time, we opt to train our own WEM. To do so, we first download the 5-grams contained in Google Books N-grams Version 2 data[10]. The corpus contains 8 million books, of which over 1.5 million are written in British English; of those, over 140,000 were written between 1958 and 1978[11]. Following Mikolov et al. (2013), we feed these texts into Google's skip-gram Word2Vec algorithm[12] whose purpose is to predict, for any given word, adjacent words in a sentence. For example, given the word 'girl,' the algorithm learns to predict nearby words like 'the,' 'wore,' 'dress': words that frequently appear in its context. To accomplish this prediction, the algorithm undertakes unsupervised learning to identify patterns in the input texts such as word co-occurrence, spacing and ordering. These patterns are encoded in a word-embedding model, which is a by-product of training the algorithm, such that each word across the texts is assigned a numerical vector.

A crucial 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

[10] This data is publicly available at: storage.googleapis.com/books/ngrams/books/datasetsv2.html.

[11] Specifically, we use the English 5-grams dataset. We train our WEM using long phrases (with up to five words) because this improves the algorithm's ability to detect text patterns. The algorithm ultimately assigns vector-values to individual words though, making it natural for us to restrict our attention to 1-grams and select 2-grams from the essays.

[12] 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).

research and simulations which find that this vector size allows for nuance in words' meanings but still provides precision (Patel and Bhattacharyya, 2017).

These 300-dimensional individual word vectors are hard to interpret, but they generate intuitive structures. For example, consider the words *king*, *man*, *woman*, *queen*, each of which is associated with the vector values (e.g. $\vec{king}$). It turns out that our WEM produces the following insight: $\vec{king} + (\vec{woman} - \vec{man}) \approx \vec{queen}$. These sorts of analogies are common to WEMs (Pennington et al., 2014) and we exploit this feature in step 5 of our procedure.

So far, we have focused on 1-grams but our Bag of Words also includes 2-grams beginning with "not" or "no." We assign such 2-grams a projection value equal to -2 times the projection value of the second word in that 2-gram, effectively reversing its gender association. For instance, we would assign the 2-gram "not marry" a projection value of -2 times that for "marry", so that negating a feminine word produces a masculine association.[13]

3.2.4 Step 4: Constructing a Gender Dimension

Kozlowski et al. (2019) argues that the reason $\vec{king} + (\vec{woman} - \vec{man}) \approx \vec{queen}$ 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 moving 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 other antonym pairs like $(\vec{female} - \vec{male})$, $(\vec{her} - \vec{him})$ etc. More generally, we can approximate Kozlowski et al. (2019)'s "gender dimension" using any pair(s) of words whose semantic difference corresponds to the feminine-masculine distinction. We draw on these arguments to construct a *gender dimension*

[13] The factor of -2 operates in two steps: the first -1 cancels out the positive (feminine) association of "marry," while the second -1 creates an equal but opposite (masculine) association for "not marry," reflecting that avoiding marriage runs counter to traditional feminine norms. We show that our results are robust to an alternative way of dealing with negations, in which we simply drop all sentences that contain the word 'no', the word 'not', or the string 'n't'.

vector, $\vec{GD}$ by taking the average of the vector values of 10 pairs of gender words:

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

We use antonym pairs (e.g., woman–man, she–he) rather than separate male and female dimensions, following established practices in the word embedding literature (Mikolov et al., 2013; Kozlowski et al., 2019). This approach addresses a fundamental measurement challenge: words like "woman" and "man" appear in contexts that relate to people generally, not just to gender. Without differencing, embedding vectors conflate gender-specific associations with broader human associations. By differencing antonym pairs, we isolate the component capturing feminine versus masculine associations, as shared human-related associations are eliminated. We further motivate and validate our gender dimension vector in Appendix C.3.

3.2.5 Step 5: Determining the gender associations of words

We now turn to measuring the gender association of 1-grams in our essays. Using the vector values we assigned to each word in step three, 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).

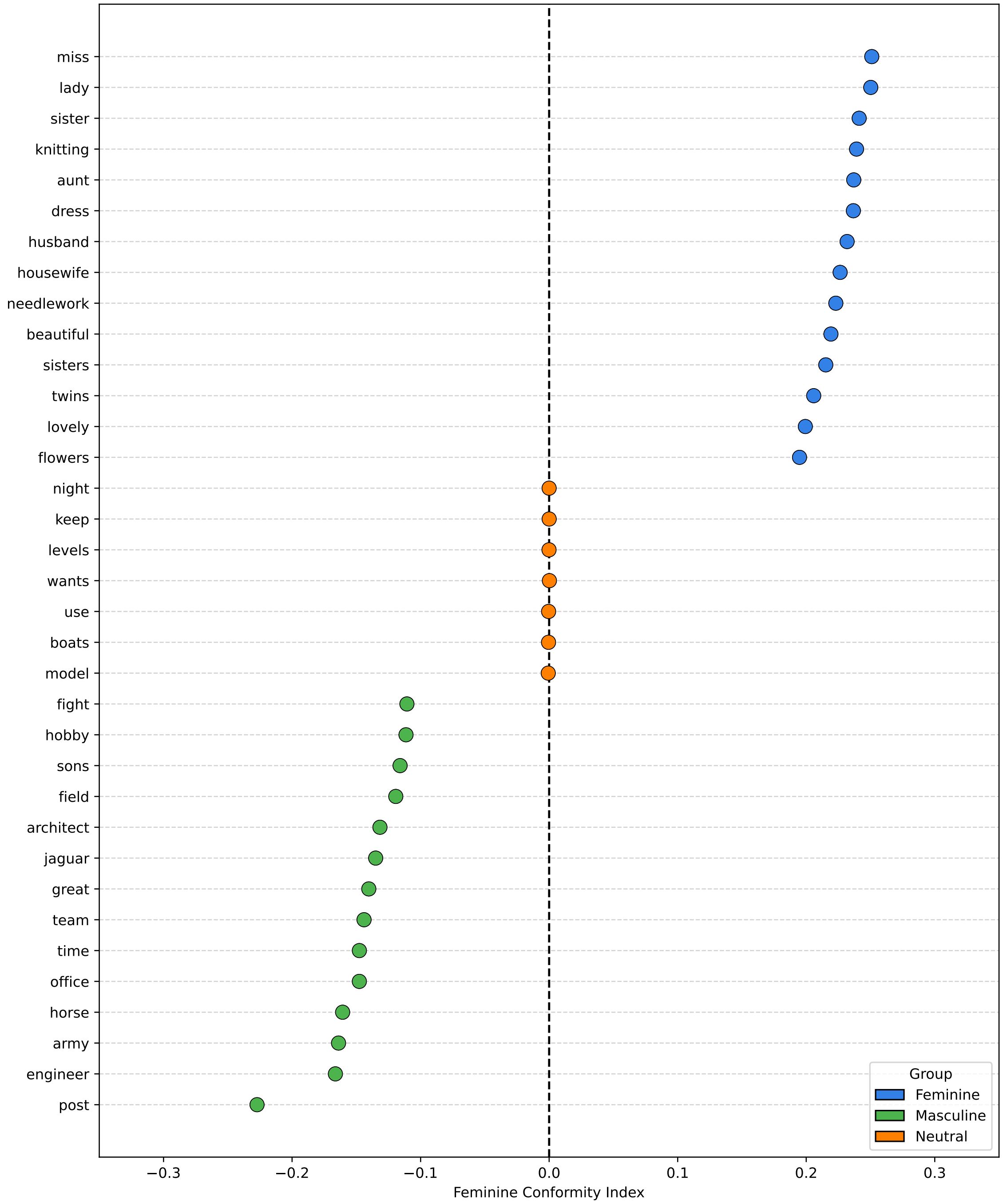
For example, to compute the gender association of "married," we project $\vec{married}$ onto $\vec{GD}$ using their dot product

$$\vec{married}^T \cdot \vec{GD} = married_1 \cdot GD_1 + married_2 \cdot GD_2 + \ldots + married_{300} \cdot GD_{300}$$

Normalizing this dot product by the lengths of both vectors yields the cosine similarity, which ranges from -1 to 1. Positive values indicate feminine associations, negative values indicate masculine associations, and larger absolute values indicate stronger associations. In our WEM, the cosine similarity of married with $\vec{GD}$ is 0.12.

To illustrate this step, Figure 2 plots projections of: (i) the 15 most feminine words (with

Figure 2: Projecting Words onto the Gender Vector



Note: The figure shows cosine similarity scores between words and the gender dimension vector, with similarity on the x-axis and words on the y-axis. 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 feminine words, green dots indicate 15 most masculine words, orange dots indicate 10 neutral words

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 judgment of the gender-connotation of the listed words, e.g. the words *dress* and *ladies* have a strong feminine association, whereas the words *army* and *team* have a strong male association.

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

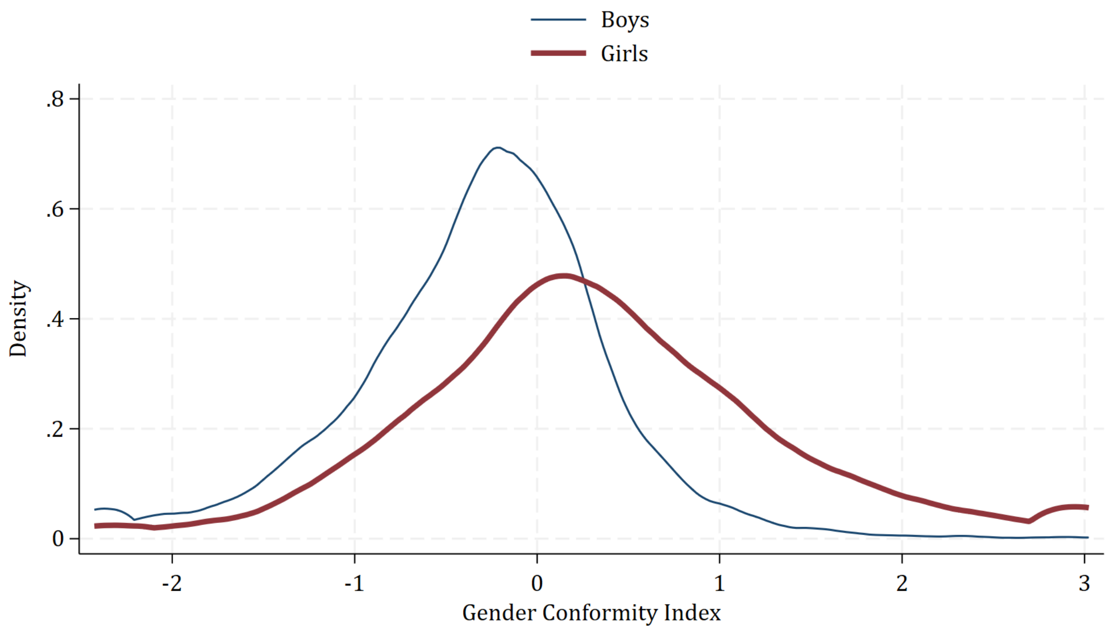
At this stage, we drop neutral words from our Bag of Words matrix to focus on words with a meaningful gender association. We do this by standardizing the projection values for all words in our Bag of Words representation and dropping those that have standardized projections less than one standard deviation away from zero. We also drop words which occur fewer than 200 times across our corpus of 10,511 essays. This is to avoid the index being driven by rarely used, or still misspelled, words. Our results are robust to both these decisions (see Section 4.4).

Finally, to move from a word-level to 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 remove the mean relationship between essay length and the weighted sum. We standardize these residuals and winsorize them at the 1st and 99th percentile to arrive at our index of feminine conformity.

3.3 Describing our Index

Figure 3 plots the distribution of our feminine conformity index for the full sample of essay writers. Girls (boys) tend to write essays that more closely conform with feminine (masculine) gender norms. The mean of the index for girls is 0.32, which is statistically different from the mean of -0.32 for boys at the 1% significance level. Interestingly, the girls' essays also display more dispersion than the boys' essays, with standard deviations of 1.01 for girls and 0.71 for boys. This is partly explained by the fact that boys use fewer gendered words overall (15.9 on average compared to 24.3 for girls), which mechanically limits the variation in their conformity scores. Moreover, boys use relatively little feminine-coded vocabulary: even the top 10% most feminine boys use only 18.3 feminine-conforming words on average, fewer than the overall average for girls. By contrast, the top 10% most feminine girls use 35.1. At the

Figure 3: Distribution of the Feminine Conformity Index for boys and girls



Notes: This graph shows the distribution of the feminine conformity index for boys and girls using the full sample of essays, that is 5,419 boys' essays and 5,091 girls essays. Higher values of the feminine conformity index indicate a greater association with feminine norms.

masculine end, the most masculine-conforming girls and boys look much more similar (31.6 and 27.9 respectively). In short, girls' essays span a wide range from masculine to highly feminine language, producing varied mixes of feminine- and masculine-conforming words across the distribution. Boys' essays, by contrast, are compressed: even the most feminine among them use little feminine-coded vocabulary, limiting the variation in conformity scores.

Drivers Words that are more "gendered" (either more masculine or feminine) or that appear more frequently in an essay, will matter most for an essay's index value. To investigate which words drive the variation in our feminine conformity index for the full sample, we use a LASSO approach. Using our feminine conformity index as the outcome and counts of all essay words weighted by their projection values as the regressors, we find that the twenty most important n-grams for predicting our index are: husband, female, time, girl, home, mother, day, wife, school, male, house, went, baby, team, job, hair, sister, tea, married. It is reassuring that many of these words might be considered gendered. See Appendix C.4.5 for more details.

Example Essay To better fix ideas on what our index captures, we present an example essay below.[14]

> "*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. [...]*"

Consistent with a casual read of this essay, which is in line with traditionally female gender roles, it has a feminine conformity index value of 0.91, which is close to two thirds of a standard deviation more feminine than the mean female essay (which equals 0.32).

Validation Checks To ensure our index is indeed capturing conformity with female gender norms, we perform several validation checks. First, we compare the projection values from our trained WEM to the ratings of gender-connotations of 360 words by 540 psychology students in a study by Jenkins et al. (1958). We find a correlation of 0.58 between our projection values and the student ratings, validating that our WEM captures gender associations in a manner consistent with human judgment from the period when the essays were written. Our method also produces a wider distribution of gender scores across words than Jenkins et al. (1958), suggesting it may capture more granular gender associations. See Appendix C.4 for details.

Next, we correlate our index of feminine conformity with leisure activities that NCDS sample members engage in at age 11. These are commonly used measures of gender conformity, shown to be different across sexes (Su et al., 2009; Banan et al., 2023). Figure A.1 reassuringly demonstrates that our index correlates positively with typically "feminine" activities – cooking, sewing, knitting, helping at home – and correlates negatively with typically "masculine" activities such as making models and playing sport. Finally, we correlate our index with three

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

questions asked at the age-16 wave, on plans for marriage and a family. We find that cohort members who write a feminine-conforming essay are more likely to want to get married and start a family earlier, and to want a larger family. These are presented in Appendix A.3.

4 Results

4.1 Lifetime Earnings and Feminine Conformity

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

$$y_i = \beta_0 + \beta_1 \text{FeminineConformityIndex}_i + \beta_2 \text{Cog11}_i + \beta_3 \text{Intern11}_i + \beta_4 \text{Extern11}_i + \gamma' \mathbf{X_i} + u_i \quad (1)$$

where y_i is one of the following: log of lifetime earnings, log of 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 each outcome, by gender. The first is a univariate regression of outcome on the feminine conformity index. The second includes cognitive and two non-cognitive skill indices (for internalizing and externalizing behavior) measured at age 11, Cog11_i, Intern11_i and Extern11_i, which are known to be important in explaining lifetime earnings and human capital accumulation (Cunha et al., 2010; Papageorge et al., 2024). The main specification includes a vector of controls, $\mathbf{X}_i$: parents' age, education and income, parents' aspirations for their child's education, number and sex composition of siblings, birth order and family stability. We also include county fixed effects in our main specification to partial out geographic variation in prevailing gender norms, thereby isolating individual-level variation in feminine conformity relative to a cohort member's local area.

Columns (1) and (4) of Table 3 report the coefficients from regressing, for girls and boys respectively, log lifetime earnings on our index of feminine conformity when not controlling for other variables. Having 1 standard deviation stronger feminine conformity is associated with a 5.4% decrease in lifetime earnings for girls, and a 4.3% decrease for boys. Put differently, an early-age tendency to conform to society's notion of "femininity" is associated with lower

Table 3: Log Lifetime Earnings

	Girls			Boys		
	Log Lifetime Earnings	Log Lifetime Earnings	Log Lifetime Earnings	Log Lifetime Earnings	Log Lifetime Earnings	Log Lifetime Earnings
Feminine Conformity Index	-0.054***	-0.038**	-0.040**	-0.043**	-0.008	-0.007
	(0.017)	(0.017)	(0.017)	(0.018)	(0.018)	(0.018)
Cognitive Skills		0.169***	0.156***		0.164***	0.143***
		(0.019)	(0.022)		(0.014)	(0.015)
Externalising Behaviour		0.050*	0.065**		-0.025	-0.020
		(0.029)	(0.030)		(0.017)	(0.017)
Internalising Behaviour		-0.045*	-0.046*		-0.041***	-0.048***
		(0.023)	(0.024)		(0.015)	(0.015)
Constant	12.581***	12.570***	12.242***	13.426***	13.431***	13.585***
	(0.018)	(0.018)	(0.353)	(0.015)	(0.015)	(0.323)
Family Background			✓			✓
County Fixed Effects			✓			✓

Notes: Sample size is 4,056 girls and 4,030 boys. We report significance at the 1% (***), 5% (**), and 10% (*) levels with robust standard errors. The first and fourth columns show the univariate OLS regression of log lifetime earnings on our feminine conformity index. Next, we include (standardized) cognitive and non-cognitive skills indices. Finally, we include (although do not display): 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. There are 111 county fixed effects. To see all coefficients displayed, see Table A.1 in Appendix A.2. All specifications include dummies for missing regressor observations.

earnings for both boys and girls, and a tendency to conform to a notion of 'masculinity' is associated with higher earnings.

Part of this association is due to the correlation of feminine conformity with cognitive skills; cognitive skills are, on average, 0.08 standard deviations lower for those with 1 standard deviation stronger conformity with female gender norms. Controlling for cognitive and non-cognitive skills, the coefficient on our index falls to (a statistically significant) 3.8% decrease for girls, and to (a statistically insignificant) 0.8% decrease for boys. Upon including parental education and income, other family background controls, and county fixed effects, the estimates are largely unchanged: a one standard deviation increase in feminine conformity predicts a statistically significant 4.0% decrease in lifetime earnings for girls and an insignificant 0.7% decrease for boys. The magnitude of this coefficient for girls is comparable to that on both non-cognitive skill indices (also measured in standard deviations), whose role in economic outcomes has been heavily studied in the literature (Cunha et al., 2010).

To understand whether these earnings results are due to wages or labor supply, Table 4 shows how feminine conformity predicts average wages, total hours worked, and share of time

Table 4: Lifetime Wages, Hours Worked

	Girls			Boys		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Log Lifetime Average Hourly Wages						
Sample mean: 1.92						
Feminine Conformity Index	-0.036***	-0.023**	-0.023**	-0.030*	0.002	-0.000
	(0.011)	(0.011)	(0.011)	(0.016)	(0.016)	(0.017)
Panel B: Log Total Lifetime Hours Worked						
Sample mean: 10.51						
Feminine Conformity Index	-0.030***	-0.016	-0.017*	-0.028***	-0.016**	-0.006
	(0.010)	(0.010)	(0.010)	(0.007)	(0.007)	(0.007)
Panel C: Share of Lifetime Spent in Employment						
Sample mean: 0.76						
Feminine Conformity Index	-0.008**	-0.002	-0.003	-0.015***	-0.008**	-0.004
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Cognitive & Non-Cognitive Skills		✓	✓		✓	✓
Parental Education & Income			✓			✓
Family Background			✓			✓
County Fixed Effects			✓			✓

Notes: Sample size is 4,056 girls and 4,030 boys. We report significance at the 1% (***), 5% (**), and 10% (*) levels with robust standard errors. Panels A and B are estimated on a sub-sample of 3,966 girls (Columns (1)-(3)) and 3,998 boys (Columns (4)-(6)) who are recorded as working in at least one month between ages 23 and 55, so wages and hours-worked are non-zero in levels. The first and fourth columns show the univariate OLS regression of each lifetime outcome on our feminine conformity index. Columns (2) and (5) includes indices for cognitive and non-cognitive skills. Columns (3) and (6) includes our full set of family background and geographic controls, including parental income and education. Family background controls include: 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 and dummies for missing regressors.

spent in employment between ages 23 and 55. Panel A reports results for the log of average hourly wages over the life cycle and Panel B reports the log of total hours worked over the life cycle, conditional on being employed at least once between ages 23 and 55. For girls, a one standard deviation increase in feminine conformity predicts a 2.3% decrease in wages a 1.7% decrease in hours worked throughout the life cycle. The latter result is consistent with Goussé et al. (2017)'s finding that gender conformity increases the time women allocate towards home production in Britain. For boys, we do not find significant associations with either total hours worked or average wages. Panel C indicates a small negative, albeit insignificant, association between our index of feminine conformity and the share of the life cycle spent in employment for boys and girls. This suggests that it is the hours of work conditional on work (or "intensive") margin that drives our hours worked results, not the employment (or "extensive") margin.

Girls versus Boys There is a sizable literature investigating the negative association between female labor-market outcomes and traditional values, preferences for home production, and beliefs on the role of women. We provide evidence that even at age 11, the extent to which young girls imagine a future that conforms with prevailing gender norms is highly predictive of lifetime earnings. On the other hand, the lack of a significant association between labor-market outcomes and feminine conformity for boys is striking. One reason for this might be that boys could face stricter social penalties for gender-nonconforming behaviors. Consistent with this conjecture, we have previously shown that boys in our sample exhibit lower variation in the feminine conformity index. This pattern aligns with emerging evidence (discussed in Matavelli et al. (2025)) that men more strictly adhere to gender norms, both in occupational choice (Pan, 2015; Delfino, 2024) and when negotiating pay (Niederle and Vesterlund, 2011; Cassar and Zhang, 2022). Given that the feminine conformity index has little predictive power for boys, we only focus on girls for the remainder of our results centered on lifetime earnings.

Lifecycle Trends Table 5 disaggregates our results on earnings, wages, and work by age. The index predicts significant earnings declines at ages 33, 42, 50, and 55. At age 33, this reflects fewer hours worked; at age 42, lower wages. These ages coincide with peak childcare responsibilities, when work interruptions lead to human capital losses and persistent earnings declines (Adda et al., 2017; Kleven et al., 2019; Gallen et al., 2023). As we show below, girls with stronger feminine conformity are more likely to have children. Thus it is unsurprising that they also have larger earnings losses at these ages. Lower hours worked persist through ages 50 and 55, suggesting lasting effects on labor supply. Table A.6 shows no clear lifecycle trends in a version of this table for boys.

4.2 Feminine Conformity Across Domains

Our feminine conformity index captures gendered expressions across all aspects of children's imagined future lives. This raises an important question: which domains of gendered expression

Table 5: Life Cycle Trends in Earnings

	Age 23	Age 33	Age 42	Age 50	Age 55
Panel A: Log Earnings in Survey Years					
Feminine Conformity Index	0.003	-0.055***	-0.065***	-0.035**	-0.048**
	(0.009)	(0.018)	(0.020)	(0.015)	(0.021)
Panel B: Log Average Hourly Wages in Survey Years					
Feminine Conformity Index	-0.001	-0.016	-0.048***	-0.000	-0.019
	(0.007)	(0.013)	(0.015)	(0.012)	(0.016)
Panel C: Log Annual Hours Worked in Survey Years					
Feminine Conformity Index	0.002	-0.018**	-0.007	-0.022***	-0.020**
	(0.004)	(0.008)	(0.007)	(0.007)	(0.008)
Panel D: Employment Status in Survey Years					
Feminine Conformity Index	0.001	-0.010	-0.001	-0.001	-0.003
	(0.008)	(0.008)	(0.007)	(0.007)	(0.009)
Cognitive & Non-Cognitive Skills	✓	✓	✓	✓	✓
Parental Education & Income	✓	✓	✓	✓	✓
Family Background	✓	✓	✓	✓	✓
County Fixed Effects	✓	✓	✓	✓	✓

Notes: Sample size is 4,056 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.

matter most for labor market outcomes? To investigate this, we decompose our overall index into domain-specific measures. Inspired by Batista and Ross (2024), we first prompt a large language model (LLM) to select the four domains most prevalent in the essays. Across all 10,511 essays, the domains it identifies are: education, work and employment ("work"), hobbies and social life ("hobbies"), and home life. We then use the LLM to categorize blocks of essay text into one of these four domains.[15] Based on this categorization, each essay is decomposed into word counts for each domain. Essentially, we have four 'sub-essays' for each individual, which we can separately score.

Table 6 presents estimates where we replace our main index with these four domain-wise indices, allowing us to identify which aspects of the essays are most predictive of lifetime labor

[15] Appendix C.6 contains the prompt we assign to the LLM to identify domains, and subsequently to categorize sentences into each domain. Reassuringly, the domains align well with the essay instructions given to the children, which features work, hobbies and home life.

market outcomes, while Appendix Table A.7 contains analogous results for boys.[16] The results reveal a dichotomy in how different domains of feminine conformity relate to labor market outcomes. Expressing stronger feminine conformity in hobbies and home life predicts lower lifetime earnings, reduced wages, fewer hours worked, and less time in employment. In contrast, expressing stronger feminine conformity in education and work predicts higher lifetime earnings, more time in employment, and higher wages, though these positive coefficients are largely statistically insignificant. This pattern suggests that the role of gender norms operates through distinct channels. Conformity to traditional feminine norms around domestic responsibilities and leisure activities reduces earnings, while conformity in professional and educational contexts may boost labor market earnings. Put differently, it is not feminine conformity per se that predicts lower earnings, but rather conformity to specifically domestic and leisure-oriented gender norms. The domains where girls imagined spending their private time – not how they imagined their careers – are what predict their actual labor market trajectories decades later.

4.3 Channels - Education, Family and Occupation

We now turn to exploring the channels that explain *why* feminine conformity predicts lower wages, labor supply, and thus earnings for girls. We examine how three key decisions – educational attainment, family formation, and occupational choice – are predicted by feminine conformity, and how they mediate our results for labor market outcomes.

Educational Attainment One potential explanation for why feminine-conforming girls earn lower wages is that they attain less education. Panel A of Table 7 supports this explanation. 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), a one standard deviation increase in gender conformity predicts 0.071 years (or about one month) less education. These declines occur across the educational distribution – feminine conformity predicts a 1.4 percentage point increase

[16] Some of the sub-essays can be empty. In these cases, we set the index to zero. To distinguish between this type of zero (no information), and a zero representing a neutral sub-essay, the regressions also contain controls for the share of each essay spent on each of the four domains. The estimated coefficients on the shares are small and statistically insignificant. Thus it is not the share of the essays devoted to different domains that is important, but the content of the words in these different domains.

Table 6: Domain-Specific Feminine Conformity Indexes (FCIs)

	Log Lifetime Earnings (1)	Share of Time Employed (2)	Log Lifetime Average Wages (3)	Log Lifetime Hours Worked (4)
FCI - Education	0.015	0.008	0.013	0.023
	(0.033)	(0.008)	(0.022)	(0.018)
FCI - Work	0.034	0.011	0.032*	0.010
	(0.025)	(0.007)	(0.017)	(0.017)
FCI - Home Life	-0.032	-0.003	-0.030**	-0.008
	(0.021)	(0.005)	(0.014)	(0.012)
FCI - Hobbies	-0.061***	-0.012**	-0.039***	-0.033**
	(0.022)	(0.005)	(0.015)	(0.013)
Cognitive & Non-Cognitive Skills	✓	✓	✓	✓
Parental Education & Income	✓	✓	✓	✓
Family Background	✓	✓	✓	✓
County Fixed Effects	✓	✓	✓	✓

Notes: Sample size is 4,056 girls. We report significance at the 1% (***), 5% (**), and 10% (*) levels with robust standard errors. Columns (3) and (4) are estimated on a sub-sample of 3,966 girls who are recorded as working in at least one month between ages 23 and 55, so wages and hours-worked are non-zero in levels. Each regression includes: indices for cognitive and non-cognitive skills and our full set of family background and geographic controls, including parental income and education, number of and sex composition of siblings, birth order, mother's and father's age at birth, parental aspiration, and family stability. Additionally, we control for essay length and the share of each essay spent on each domain. We include 111 county fixed effects.

in high school dropout rates (relative to a mean of 27 percent) and a 1.1 percentage point decline in university attendance (relative to a mean of 24 percent). Among girls who completed high school, a one standard deviation increase in gender conformity predicts a 1.4 percentage point decline in university attendance, suggesting that feminine conformity influences multiple educational decisions.[17]

Family Formation We explore two family formation decisions, marriage and fertility, which could be influenced by feminine conformity and could influence earnings. The first two columns of Panel B in Table 7 show regression estimates for whether the cohort member is married and has a child at 23. The latter two columns consider whether she has *ever* married and

[17] Panel A tests four hypotheses. To test whether the four are jointly significant, accounting for issues of multiple hypothesis testing, we use the approach in Viviano et al. (2025). In particular, we construct an educational index that is a weighted average of the four estimates in panel A. We find that a one standard deviation increase in feminine conformity predicts a decrease in the education index that is statistically significant at the 5% level. Similarly, using the results in Panel B, we construct a fertility index that has a positive association with our index, also significant at the 5% level. However, a marriage index created using the two marriage results does not exhibit a statistically significant association with feminine conformity.

Table 7: Predicting Education, Family Formation and Occupation

	(1) Years of Education	(2) HS Dropout	(3) Attend Uni	(4) Attend Uni if HS completed
Panel A: Education				
Model:	OLS	Logit	Logit	Logit
Feminine Conformity Index	-0.071***	0.014**	-0.011*	-0.014*
	(0.022)	(0.006)	(0.006)	(0.008)
Sample Mean:	11.99	0.27	0.24	0.33
	Married at 23	**Any Children at Age 23**	**Ever Married**	**Ever Have Children**
Panel B: Family Formation				
Model:	Logit	Logit	Logit	Logit
Feminine Conformity Index	0.011	0.004	0.002	0.014**
	(0.008)	(0.007)	(0.006)	(0.007)
Sample Mean:	0.56	0.33	0.86	0.83
	Professional	**Administrative**	**Manual**	**Elementary**
Panel C: Occupation				
Model:	Multi-Logit	Multi-Logit	Multi-Logit	Multi-Logit
Feminine Conformity Index	-0.012*	-0.005	0.011	0.006
	(0.006)	(0.008)	(0.008)	(0.006)
Sample Mean:	0.20	0.35	0.29	0.16
Cognitive & Non-Cognitive Skills	✓	✓	✓	✓
Parental Education & Income	✓	✓	✓	✓
Family Background	✓	✓	✓	✓
County Fixed Effects	✓	✓	✓	✓

Notes: Sample size is 4,056 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 feminine 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 controlling 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 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. Qualitative results in Panels A and B are not sensitive to using a probit or linear probability model, instead of logit, for the binary outcome variables.

had children. We find that a one standard deviation increase in feminine conformity predicts a significant 1.4 percentage point increase in the probability of ever having a child (in a sample where 83% of women have children). There is a large literature discussing the cost of lost career earnings from having children (Adda et al., 2017; Kleven et al., 2019; Gallen et al., 2023). However, we find no statistically significant evidence that feminine conformity is associated with the timing of family formation or ever being married.

Occupational Choice Finally, we consider whether feminine conformity predicts occupational choice for girls, which could have implications for both labor supply and wages. Girls with one standard deviation stronger feminine conformity are 1.2 percentage points less likely to enter professional occupations.[18] Since professional occupations pay more than manual ones[19], the lower earnings predicted by feminine conformity could in part be explained by differences in occupation choice.

Mediation Analysis Table 8 quantifies the extent to which educational attainment, family formation and occupational choice mediate our lifetime earnings results for girls. We use a Gelbach (2016) decomposition to show the estimated contribution of each of these three decisions to the coefficient on feminine conformity. The Gelbach decomposition is robust to the sequence in which they are included. Details on the decomposition can be found in Appendix D.

Table 8: Share of Association Mediated through Education, Family and Occupational Choice

	Coefficient		Percent Explained by Mediator			
	Baseline	With Mediators	Education	Family Formation	Occupation	Total
Log Lifetime Earnings	-0.040	-0.025	11.1%	13.4%	17.0%	41.6%
Log Avg Wage	-0.023	-0.012	23.0%	14.0%	15.7%	52.7%
Log Total Hours Worked	-0.017	-0.008	9.0%	20.7%	15.4%	45.1%
Employment	-0.003	-0.000	23.3%	36.9%	17.0%	77.2%

Notes: Sample size is 4,056 girls. Rows (2) and (3) ("Log Avg Wage" and "Log Total Hours Worked") are estimated on a sub-sample of 3,966 girls who are recorded as working in at least one month between ages 23 and 55, so wages and hours-worked are non-zero in levels. 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 (ever married and total number of children), 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.

[18] Professional occupations include: natural scientists, engineers, technologists, health professionals, teaching professionals, legal professionals, business and financial professionals, architects, town planners, and librarians. This is based on the UK's Office for National Statistics' Standard Occupational Classification 2000.

[19] 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.

The first column ("Baseline") in Table 8 contains the coefficients on gender conformity from our main specifications reported in the third columns of Tables 3 (for earnings) and 4 (for wages, hours worked and employment). The second column reports the coefficients on gender conformity from regressions which also include our mediators: education, family formation, 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 total shares of the baseline coefficients that are explained by these mediators.[20]

Comparing estimates from the models with and without mediators reveals that 42% of the association between feminine conformity and lifetime earnings can be accounted for by education, family formation, and occupation. Furthermore, our mediators explain 53%, 45%, and 77% of feminine conformity's association with wages, hours, and employment, respectively. Education and occupation account for more of the earnings and the wage results whereas family formation and occupation account for more of the working hours results. To summarize: women with stronger feminine conformity attain less education, which reduces their wages, and are likely to have children, which reduces their working hours. Occupations chosen by those with stronger feminine conformity predict lower wages and fewer hours worked.

4.4 Robustness Checks

In this section, we perform several robustness checks for the results for girls in Tables 3 and 4. We address concerns related to omitted variables, regression misspecification, construction of feminine conformity or lifetime earnings and measurement error in the index.

Other Latent Information Contained in the Essays A natural concern is that our index captures unobservables correlated with both earnings and feminine conformity, such as educational or occupational aspirations. Following Kozlowski et al. (2019), we construct six additional indices using the same NLP methods: (1) education (educated-uneducated), (2) cultivation (cultured-uncultured), (3) affluence (rich-poor), (4) status (prestigious-undistinguished),

[20]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.

Table 9: Alternative Explanations

	Lifetime Earnings	Lifetime Average Wages	Employed	Log Total Hours Worked
Panel A: Other Text Factors				
Feminine Conformity Index	-0.037**	-0.024**	-0.001	-0.016
	(0.018)	(0.012)	(0.004)	(0.011)
Education Score	-0.012	0.000	-0.000	-0.011
	(0.019)	(0.013)	(0.005)	(0.013)
Cultivation Score	0.026	0.004	0.001	-0.001
	(0.024)	(0.016)	(0.006)	(0.016)
Affluence Score	0.019	0.034***	-0.006	-0.005
	(0.019)	(0.013)	(0.005)	(0.012)
Status Score	-0.010	-0.005	-0.004	-0.002
	(0.018)	(0.012)	(0.004)	(0.011)
Morality Score	-0.014	-0.023*	0.005	0.001
	(0.020)	(0.013)	(0.005)	(0.012)
Employment Score	-0.018	-0.015	-0.003	-0.010
	(0.017)	(0.011)	(0.004)	(0.010)
Panel B: Complementarities Between Covariates – Post-LASSO Model				
Feminine Conformity Index	-0.041**	-0.025**	-0.003	-0.018*
	(0.017)	(0.011)	(0.004)	(0.011)
Panel C: Removing Negation Phrases				
Feminine Conformity Index	-0.048***	-0.025**	-0.004	-0.021**
	(0.017)	(0.011)	(0.004)	(0.010)
Cognitive & Non-Cognitive Skills	✓	✓	✓	✓
Parental Education & Income	✓	✓	✓	✓
Family Background	✓	✓	✓	✓
County Fixed Effects	✓	✓	✓	✓

Notes: Sample size is 4,056 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 coefficient from a post-LASSO regression – we only report the index, which is always selected by LASSO (unlike non-cognitive skills). We add dummies for missing regressor observations and fixed effects to the post-LASSO specification.

(5) morality (good-evil), and (6) employment (employer-employee).[21] Panel A of Table 9 shows our main results are robust to including these six indices in the regression. One standard devia-

[21]For each dimension, we follow steps 1-3 from Section 3.2, then construct dimension vectors using relevant antonym pairs (e.g., educated-uneducated, learned-unlearned for education). We project essay words onto each dimension vector and aggregate weighted word counts as in step 5. See Appendix C.6 for details.

tion more feminine conformity predicts a 3.7% decrease in lifetime earnings, a 2.4% decrease in average wages, and a 1.6% decrease in hours, which are all close to the baseline values reported in Tables 3 and 4.

Complementarities between Predictors We next address concerns that our results may be driven by complementarities and interactions between our control variables. 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. First, our post-LASSO regressions always select our index. Panel B in Table 9 shows that this leads to very similar declines in earnings, wages, and labor supply predicted by feminine conformity as in our baseline.

Removing Negation Phrases Our baseline feminine conformity index is constructed using all 1-grams as well as 2-grams starting with "no" and "not". However, this does not comprehensively remove all negation phrases. For example, a girl might say "All my friends will be married, but I won't be." Our index would include a score for the word "married", despite the writer actually not wanting to get married. We check robustness to this issue by removing all sentences with "no", "not" and words with a suffix of "n't" (e.g. can't, won't). This reduces mean essay length for girls from 228 words to 191 words. Panel C in Table 9 shows this more conservative approach to dealing with negations leads to larger estimated declines in earnings, wages and hours worked over the life cycle.

Additional Controls for Skills, School Quality or Aspirations Our index could be proxying for other important predictors of lifetime outcomes. In Table A.10, 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, (iv) the occupation the girl declares to aspire to in the survey questionnaire and (v) region (instead of county) fixed effects.[22] In each case, our results are very similar to those reported in Tables 3 and 4.

[22]The county fixed effects are added to address the concern that the content of the essays may reflect the gender norms of those living in the child's county. If so, the relationship between the feminine conformity index

Index Construction We explore robustness of our lifetime earnings result to two key components of how we estimate our index of feminine conformity. First, we re-construct the index using alternative word-embedding models (WEMs). We train five alternative WEMs using text from five non-overlapping periods – 1955-1959, 1960-1964, 1965-1969, 1970-1974 and 1975-1979. These time periods reflect alternative windows in the period around when the essays were written (in 1969).[23]

Table A.11 shows that our results do not qualitatively change and remain largely statistically significant when we alter the specific time window for texts that we use to train our WEM. Interestingly, estimates are larger when using WEMs trained using text from before the essays were written (1955-1964) than when using text after when the essays were written (1970-1979). The WEM trained using text written during the time the essays were written (1965-1969) produces the same earnings estimate as our baseline estimate of -0.040.[24] Next, 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.12 confirms that our results are mostly robust to these choices.

Construction of Lifetime Earnings For individuals who never worked between ages 23-55, we set lifetime earnings to the first percentile of earnings for each gender, allowing those individuals to remain in the sample when we take logarithms. Table A.13 shows that our headline result is robust to using alternative approaches, including using earnings in levels and dropping zero-earners. For example, when we drop those who never work, a one standard

and later life outcomes might only capture county level differences in outcomes. To better understand the extent to which county level variation may be important, we take the mean value of the index of everyone in i's county (except i), and correlate this with i's value of the index. By eliminating i from the mean value of the index of everyone in i's county, we eliminate the mechanical correlation the two variables. This correlation is 0.05, with each county having 75 cohort members on average. This correlation is low, highlighting that the great majority of variation in our index is not explained by local area factors.

[23] Our baseline index uses texts between 1958 and 1978, a 10 year window either side of the essays. We cover the period from birth of the NCDS cohort (in 1958) to 10 years after the essays were written using five non-overlapping time windows.

[24] We also explore robustness to using two pre-trained word-embedding models trained on extensive corpora of predominantly modern, digitized texts; specifically, Google's Word2Vec and Stanford's GloVe WEMs. However, when we use these WEMs, the results are attenuated. This is consistent with Kozlowski et al. (2019)'s finding that gender connotations in English have evolved throughout the twentieth century. See Appendix C.2.1 for a detailed discussion.

deviation increase in the feminine conformity index decreases lifetime earnings by 3.4%. This is consistent with our previous findings that while our results are mostly driven by wages and the intensive margin of labor supply, there is a modest extensive labor supply margin at work also. The table also tests alternative imputation methods to address missing earnings data (due to non-response) when constructing lifetime earnings. Our headline result is robust across all specifications.

4.5 Quality of Life and Health

As shown previously, the feminine conformity index predicts educational attainment, family formation, and labor market outcomes, all of which are likely to impact later-life wellbeing. Here we investigate whether the feminine conformity index is associated with quality of life and health at age 55.

Panel A of Table 10 reports results for the (standardized) CASP-12 quality of life score, which captures four aspects of quality of life – control, autonomy, self-realization, and pleasure (Hyde et al. (2003); see Appendix B.5 for details). Columns (1) and (4) report the coefficients from regressing quality of life on our feminine conformity index when not controlling for other variables, whereas columns (2) and (5) also control for skills and columns (3) and (6) control for the full set of covariates. For both boys and girls, a 1 standard deviation increase in feminine conformity is associated with an approximately 0.04 standard deviation decline in the CASP-12 score; that is, with a lower quality of life.

Feminine conformity relates to several outcomes that affect quality of life. We next examine whether the association between feminine conformity and quality of life operates through the earnings channel we documented earlier. If feminine conformity reduces quality of life primarily through its association with lower lifetime earnings, then controlling for earnings should substantially reduce the estimated relationship. However, when we add lifetime earnings in the regressions, the estimated relationship between quality of life and feminine conformity changes only modestly. The coefficient for girls falls from -0.043 in column (3) to -0.039 when earnings are included, while the coefficient for boys rises to -0.047.[25]

Panels B and C report results for indices of mental and physical health respectively. A one standard deviation higher feminine conformity predicts 0.033 standard deviations lower mental health for girls, and 0.052 lower for boys. In other words, feminine conformity is associated with worse mental health later in life. As with the quality of life results in Panel A, including lifetime earnings (and indeed other mediators) only modestly attenuates these coefficients, suggesting that feminine conformity directly predicts lower mental health not only indirectly via worse labor market outcomes. Our data do not allow us to account for why we find these patterns. For boys, this might be due to the social penalties of deviating from the norm, or the personal cost of having to fit into a masculine role (Sarzosa and Urzúa, 2021; Banan et al., 2023). For girls this might be that gender norms constrain women's decisions (unrelated to the labor market), which worsens their mental wellbeing (Heise et al. (2019)). For boys and girls, we find no evidence that our feminine conformity index predicts physical health.

4.6 Predictors of Feminine Conformity

What determines conformity with gender norms? Recent studies show that attitudes towards gender norms are malleable (Dhar et al., 2022; Bursztyn et al., 2020), and that family background characteristics, such as sibling gender composition (Brenøe, 2022) and parental characteristics (Fernández et al. (2004), Dhar et al. (2019)), affect conformity with gender norms. Our setting is a particularly interesting one in that we can examine the micro-determinants of feminine conformity among children who experience relatively homogeneous upbringings (i.e., all born in Britain in 1958).

We regress our feminine conformity index on a large set of potential determinants, including child skills, family demographics, role model variables, and region-level characteristics. Figure 4 reports coefficients from a saturated regression in which we include all determinants, for boys and girls separately. Table A.2 displays coefficients from univariate and saturated regressions.

[25] When we additionally control for our mediators from Section 4.3 – educational attainment, family formation and occupation – the coefficients drop to -0.038 and -0.043 for girls and boys respectively, significant at the 5% level for girls, although insignificant for boys.

Table 10: Quality of Life and Health

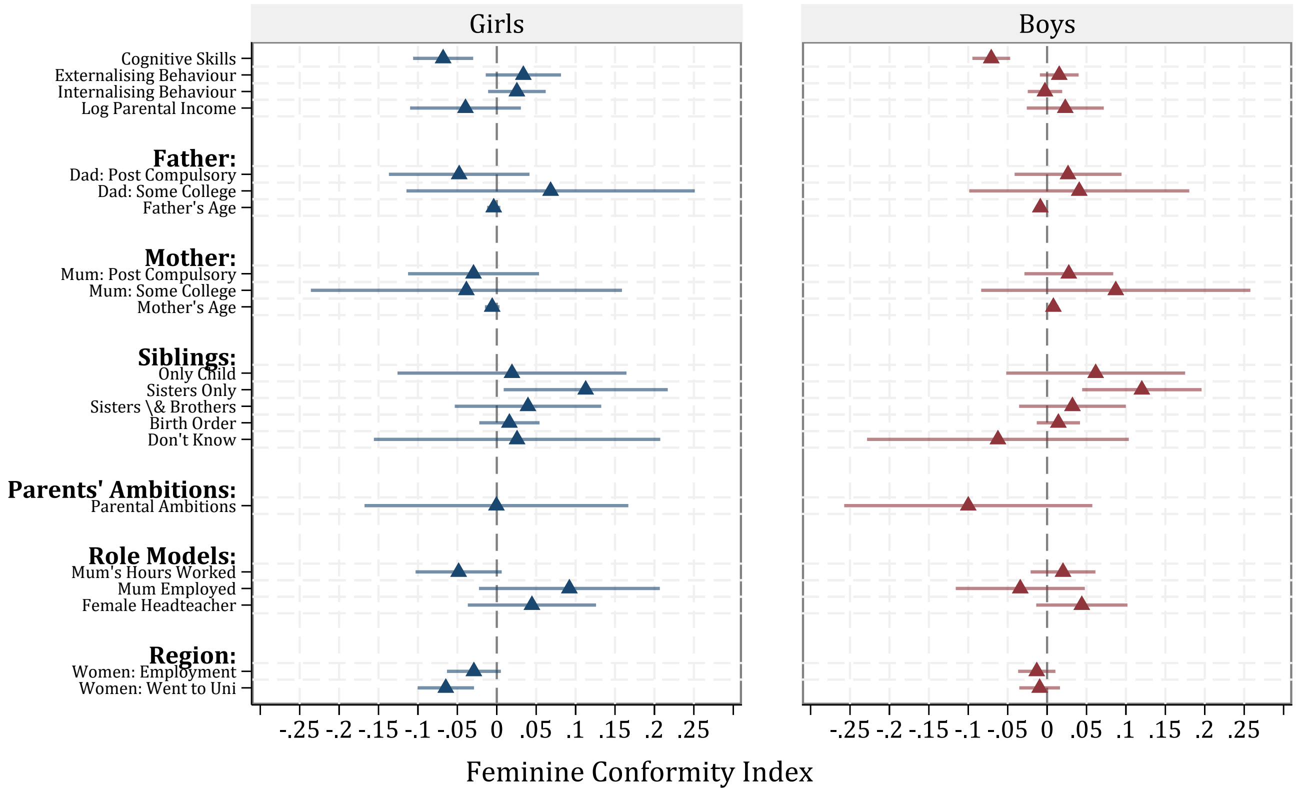
	Girls			Boys		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Quality of Life						
Feminine Conformity Index	-0.050***	-0.039**	-0.043**	-0.057**	-0.046*	-0.046*
	(0.018)	(0.018)	(0.019)	(0.026)	(0.026)	(0.026)
Panel B: Mental Health						
Feminine Conformity Index	-0.038**	-0.025	-0.033*	-0.052**	-0.046*	-0.052*
	(0.019)	(0.019)	(0.020)	(0.025)	(0.025)	(0.027)
Panel C: Physical Health						
Feminine Conformity Index	-0.013	0.011	0.007	-0.045*	-0.020	-0.021
	(0.017)	(0.017)	(0.018)	(0.024)	(0.024)	(0.025)
Cognitive & Non-Cognitive Skills		✓	✓		✓	✓
Controls			✓			✓
County FE			✓			✓
Parental Educ/Inc			✓			✓

Notes: Sample size is 4,056 girls and 4,030 boys. We report significance at the 1% (***), 5% (**), and 10% (*) levels with robust standard errors. The first and fourth columns show the univariate OLS regression of each outcome on our feminine conformity index. Columns (2) and (5) includes indices for cognitive and non-cognitive skills. Columns (3) and (6) includes our full set of family background and geographic controls, including parental income and education. Family background controls include: 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 regressors.

Three insights stand out. First, higher age-11 cognitive skills are associated with lower feminine conformity. Controlling for the full set of covariates, a one standard deviation improvement in cognitive skills is associated with a 0.07 standard deviation decline in feminine conformity for girls, and a 0.08 standard deviation decline for boys. Next, we find evidence that, relative to having only brothers (who are the reference group), girls and boys who have only sisters tend to display 0.1 standard deviations higher feminine conformity. This is in line with previous studies that find that the composition of siblings can influence gender conformity by affecting gender roles within the household (we find that having only sisters predicts higher feminine conformity whereas Brenøe (2022) find having brothers does).

Finally, we find that, for girls (but not for boys), lower feminine conformity is associated with growing up in an area with a higher fraction of women who attend university relative to men and a higher fraction of working women. This is consistent with Fernández (2007),

Figure 4: Determinants of Feminine Conformity Index



Notes: Sample is 4,056 girls and 4,030 boys. Triangles represent point estimates, bars represent 95% confidence intervals, computed using robust standard errors. The coefficients are from from a regression of the feminine conformity index on all the variables shown in the figure, plus indicators for missing observations. The graph on the left is for girls, on the right is boys. Table A.3 in the appendix shows summary statistics for all determinants.

who find that women migrating from countries with traditional gender norms continue to make decisions consistent with those traditional norms, and Farré and Vella (2013) who find evidence of norms persisting across generations. One interpretation of this result is that in areas where more women attend university or work, misperceptions of social norms (in particular, believing society is more gender conservative than it is) are less prevalent (Cortés et al., 2024).

The R^2 values from the regressions in Figure 4 are 0.024 for girls and 0.030 for boys. The Shapely values reported in Tables A.4 and A.5 show that cognitive skills explain by far the highest share of the R^2 in the models for both boys and girls. This is followed by the fraction of women who attend university and non-cognitive skills (for girls). Given our previous results, an interesting avenue for future work would be to understand what else drives gender conformity.[26]

4.7 Interpreting our Index

Our empirical results show that age 11 gender conformity predicts women's labor market outcomes, but not men's. However, we have so far been agnostic about *why* gender conformity predicts these outcomes. Appendix E presents a simple model to show two potential interpretations of why gender conformity may affect labor market outcomes.

One interpretation of our feminine conformity index is that it measures a child's preference over engaging in typically female activities and behaviors. For example, our index may capture heterogeneity in utility derived from having children or from home-produced goods. A second interpretation is that our index measures constraints faced by women (either real or perceived by them as young children). For instance, some girls might perceive that their labor market return to education is low, prompting them to engage in more traditionally female activities, despite not preferring to do so.

In our model, both of these interpretations imply that feminine-conforming girls – that is, girls with a stronger preference for home-produced goods or a lower perceived return to education – attain less education. This is consistent with the evidence shown in Table 7.

If higher gender conformity captures stronger preferences for home production, our model provides insights on labor supply which match our data. In particular, Table 4 shows that women with higher gender conformity work fewer hours over their life cycle. Relatedly, Table 7 shows that our index predicts higher marriage and child-bearing rates, corresponding to a larger share of adult life in home production. Table 6 also appears to support this interpretation. It shows that a higher degree of female conformity in hobbies and home life – the domains most closely tied to preferences over domestic and leisure activities – are particularly predictive of lower lifetime earnings.

An open question that remains is why feminine conformity does not predict labor market

[26] For example, Kågesten et al. (2016) review the literature in psychology, sociology, and gender studies on the formation of gender attitudes in adolescents. They highlight that school characteristics, media, and the influence of friends may be important determinants.

outcomes for boys. Our model suggests that preferences towards household specialization results in greater home production and reduced labor supply for women. However, men are often less likely to be close to this decision margin, tending to allocate more of their time to work outside the household (Bolt et al., 2023; Burda et al., 2013; Blundell et al., 2016).

Ultimately, even though we cannot conclusively identify whether preferences or constraints drive the associations we observe, our empirical findings add to our understanding of women's human capital investments and labor supply decisions. If our results arise due to preference heterogeneity, 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 index. If instead the observed patterns reflect real or perceived constraints, then our results point to the importance of better understanding how these constraints arise. We leave it for future research to further disentangle the role of these two potential mechanisms in generating our results.

5 Conclusion

In this paper, we use natural language processing and novel text data from thousands of essays written by 11-year-old children born in Britain in 1958 to construct an index of gender conformity. 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 labor market outcomes. Our approach has several important strengths. First, we observe gender conformity in childhood, meaning that concerns of justification bias (due to labor market outcomes affecting gender attitudes) are minimal. Second, by training our own word embedding model on over 140,000 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 panel data that tracks individuals from childhood to 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. 40% of the association between gender conformity and earnings, and over half of the association with wages 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. Indeed, while our findings focus on feminine conformity for girls in 1969 Britain, the evolving nature of gender norms suggests that the role of masculine conformity in shaping boys' outcomes is also a promising area for future research.

Recent advances using Large Language Models will allow for additional insights. For example, while we focus on gender conformity, there are other dimensions of people's aspirations that could be studied, including family, career, and educational attainment. Newer methods that use not just counts of different words, but also the phrasing of those words, might be particularly valuable in understanding people's aspirations more broadly.

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 (2016). 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.

Agostinelli, F. and M. Wiswall (2025). Estimating the technology of children's skill formation. Journal of Political Economy 133(3), 846–887.

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.

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.

Batista, R. M. and J. Ross (2024). Words that work: Using language to generate hypotheses. Available at SSRN 4926398.

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

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 (2007). Accounting for intergenerational income persistence: noncognitive skills, ability and education. The Economic Journal 117(519), C43–C60.

Blundell, R., L. Pistaferri, and I. Saporta-Eksten (2016). Consumption inequality and family labor supply. American Economic Review 106(2), 387–435.

Boelmann, B., A. Raute, and U. Schönberg (2025). Wind of change? cultural determinants of maternal labor supply. American Economic Journal: Applied Economics 17(2), 41–74.

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

Bolt, U., E. French, J. Hentall-MacCuish, and C. O'Dea (2023). Intergenerational altruism and transfers of time and money: a life cycle perspective. Technical report, IFS Working Papers.

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.
Brazier, J. E., R. Harper, N. Jones, A. O'Cathain, K. J. Thomas, T. Usherwood, and L. Westlake (1992). Validating the sf-36 health survey questionnaire: new outcome measure for primary care. British medical journal 305(6846), 160–164.
Brenøe, A. A. (2022). Brothers Increase Women's Gender Conformity. Journal of Population Economics 35(4), 1859–1896.
Burda, M., D. S. Hamermesh, and P. Weil (2013). Total work and gender: facts and possible explanations. Journal of population economics 26(1), 239–261.
Bursztyn, L., T. Fujiwara, and A. Pallais (2017). 'acting wife': Marriage market incentives and labor market investments. American Economic Review 107(11), 3288–3319.
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.
Cassar, A. and Y. J. Zhang (2022). The competitive woman: Evolutionary insights and cross-cultural evidence into finding the femina economica. Journal of Economic Behavior & Organization 197, 447–471.
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.
Clarke, A., T. Friede, R. Putz, J. Ashdown, S. Martin, A. Blake, Y. Adi, J. Parkinson, P. Flynn, S. Platt, et al. (2011). Warwick-edinburgh mental well-being scale (wemwbs): validated for teenage school students in england and scotland. a mixed methods assessment. BMC public health 11(1), 487.
Cortés, P., G. Koşar, J. Pan, and B. Zafar (2024). Should mothers work? how perceptions of the social norm affect individual attitudes toward work in the us. Review of Economics and Statistics, 1–28.
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.
Delfino, A. (2024). Breaking gender barriers: Experimental evidence on men in pink-collar jobs. American Economic Review 114(6), 1816–1853.
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.

Elliott, J. (2010). Imagining a gendered future: Children's essays from the national child development study in 1969. Sociology 44(6), 1073–1090.

Farré, L. and F. Vella (2013). The intergenerational transmission of gender role attitudes and its implications for female labour force participation. Economica 80(318), 219–247.

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.

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

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.

Gallen, Y., J. S. Joensen, E. R. Johansen, and G. F. Veramendi (2023). The labor market returns to delaying pregnancy. Available at SSRN 4554407.

Gauri, V., T. Rahman, and I. K. Sen (2019). Measuring social norms about female labor force participation in jordan. World Bank Policy Research Working Paper (8916).

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.

Ghodsian, M. (1977). Children's behaviour and the bsag: some theoretical and statistical considerations. British Journal of Social and Clinical Psychology 16(1), 23–28.

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.

Hamilton, W. L., J. Leskovec, and D. Jurafsky (2016). Diachronic word embeddings reveal statistical laws of semantic change. arXiv preprint arXiv:1605.09096.

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.

Hays, R. D., C. D. Sherbourne, and R. M. Mazel (1993). The rand 36-item health survey 1.0. Health economics 2(3), 217–227.

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.

Heise, L., M. E. Greene, N. Opper, M. Stavropoulou, C. Harper, M. Nascimento, D. Zewdie, G. L. Darmstadt, M. E. Greene, S. Hawkes, et al. (2019). Gender inequality and restrictive gender norms: framing the challenges to health. The Lancet 393(10189), 2440–2454.

Hyde, M., R. D. Wiggins, P. Higgs, and D. B. Blane (2003). A measure of quality of life in early old age: the theory, development and properties of a needs satisfaction model (casp-19). Aging & mental health 7(3), 186–194.

Ichino, A., M. Olsson, B. Petrongolo, and P. Skogma-Thoursie (2026). Taxes, childcare and gender identity norms. Journal of Labor Economics.

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.

Kågesten, A., S. Gibbs, R. W. Blum, C. Moreau, V. Chandra-Mouli, A. Herbert, and A. Amin (2016). Understanding factors that shape gender attitudes in early adolescence globally: A mixed-methods systematic review. PloS one 11(6), e0157805.

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. Technical report. Unpublished.

Kleven, H. and C. Landais (2017). Gender inequality and economic development: fertility, education and norms. Economica 84(334), 180–209.

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.

Lin, Y., J.-B. Michel, E. A. Lieberman, J. Orwant, W. Brockman, and S. Petrov (2012). Syntactic annotations for the google books ngram corpus. In Proceedings of the ACL 2012 system demonstrations, pp. 169–174.

Matavelli, I., P. Grosjean, R. De Haas, and V. Baranov (2025). Masculinity norms and their economic consequences. Technical report, Working Paper.

McHorney, C. A., J. E. Ware Jr, and A. E. Raczek (1993). The mos 36-item short-form health survey (sf-36): Ii. psychometric and clinical tests of validity in measuring physical and mental health constructs. Medical care, 247–263.

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

Michel, J.-B., Y. K. Shen, A. P. Aiden, A. Veres, M. K. Gray, G. B. Team, J. P. Pickett, D. Hoiberg, D. Clancy, P. Norvig, et al. (2011). Quantitative analysis of culture using millions of digitized books. Science 331(6014), 176–182.

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.

Ng Fat, L., S. Scholes, S. Boniface, J. Mindell, and S. Stewart-Brown (2017). Evaluating and establishing national norms for mental wellbeing using the short warwick–edinburgh mental well-being scale (swemwbs): findings from the health survey for england. Quality of Life Research 26(5), 1129–1144.

Niederle, M. and L. Vesterlund (2011). Gender and competition. Annu. Rev. Econ. 3(1), 601–630.

Olivetti, C., J. Pan, and B. Petrongolo (2024). The evolution of gender in the labor market. In Handbook of labor economics, Volume 5, pp. 619–677. Elsevier.

Pan, J. (2015). Gender segregation in occupations: The role of tipping and social interactions. Journal of Labor Economics 33(2), 365–408.

Papageorge, N. W., V. Ronda, and Y. Zheng (2024). The economic value of breaking bad: Misbehavior, schooling and the labor market.

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.

Pechenick, E. A., C. M. Danforth, and P. S. Dodds (2015). Characterizing the google books corpus: Strong limits to inferences of socio-cultural and linguistic evolution. PloS one 10(10), e0137041.

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.

Richardson, K., M. Calnan, J. Essen, and L. Lambert (1976). The linguistic maturity of 11-year-olds: Some analysis of the written compositions of children in the national child development study. Journal of Child Language 3(1), 99–115.

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.

Sarzosa, M. and S. Urzúa (2021). Bullying among adolescents: The role of skills. Quantitative economics 12(3), 945–980.

Stewart-Brown, S., A. Tennant, R. Tennant, S. Platt, J. Parkinson, and S. Weich (2009). Internal construct validity of the warwick-edinburgh mental well-being scale (wemwbs): a rasch analysis using data from the scottish health education population survey. Health and

quality of life outcomes 7(1), 15.
Su, R., J. Rounds, and P. I. Armstrong (2009). Men and things, women and people: a meta-analysis of sex differences in interests. Psychological bulletin 135(6), 859.
Tennant, R., L. Hiller, R. Fishwick, S. Platt, S. Joseph, S. Weich, J. Parkinson, J. Secker, and S. Stewart-Brown (2007). The warwick-edinburgh mental well-being scale (wemwbs): development and uk validation. Health and Quality of life Outcomes 5(1), 63.
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.
Viviano, D., K. Wuthrich, and P. Niehaus (2025). A model of multiple hypothesis testing. arXiv preprint arXiv:2104.13367.
Walters, S. J., J. F. Munro, and J. E. Brazier (2001). Using the sf-36 with older adults: a cross-sectional community-based survey. Age and ageing 30(4), 337–343.
Ware Jr, J. E. (2000). Sf-36 health survey update. Spine 25(24), 3130–3139.
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.
Wiggins, R. D., G. Netuveli, M. Hyde, P. Higgs, and D. Blane (2008). The evaluation of a self-enumerated scale of quality of life (casp-19) in the context of research on ageing: A combination of exploratory and confirmatory approaches. Social Indicators Research 89(1), 61–77.
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.

Appendix for Online Publication

A Additional Results

A.1 Additional Tables

Lifetime Earnings - All Coefficients

Table A.1: Lifetime Earnings - All Coefficients

	Girls Log Lifetime Earnings	**Boys** Log Lifetime Earnings
Feminine Conformity Index	-0.040**	-0.007
	(0.017)	(0.018)
Cognitive Skills	0.156***	0.143***
	(0.022)	(0.015)
Externalising Behaviour	0.065**	-0.020
	(0.030)	(0.017)
Internalising Behaviour	-0.046*	-0.048***
	(0.024)	(0.015)
Dad: Post Compulsory	0.058	0.010
	(0.048)	(0.037)
Dad: Some College	-0.001	0.109
	(0.093)	(0.079)
Mum: Post Compulsory	-0.018	0.026
	(0.044)	(0.033)
Mum: Some College	0.193	-0.019
	(0.121)	(0.102)
Log Parental Income	-0.009	0.016
	(0.040)	(0.032)
Only Child	0.125	0.032
	(0.079)	(0.062)
Sisters Only	0.066	-0.047
	(0.056)	(0.043)
Sisters & Brothers	0.028	-0.027
	(0.053)	(0.040)
Number of Siblings	0.014	0.007
	(0.016)	(0.012)
Birth Order	-0.022	-0.029
	(0.023)	(0.018)
Expect to go to School	0.046	-0.034
	(0.109)	(0.079)
Expect to go to Uni	0.051	0.012
	(0.131)	(0.122)
Mother's Age	-0.002	0.002
	(0.005)	(0.004)
Father's Age	0.005	0.001
	(0.004)	(0.003)
Stable Family	0.038	0.045
	(0.054)	(0.044)
County Fixed Effects	✓	✓

Notes: Sample size is 4,056 girls and 4,030 boys. We report significance at the 1% (***), 5% (**), and 10% (*) levels with robust standard errors. In this table, we present all coefficients from the specifications in Columns 3 and 6 of Table 3. We also include 111 county fixed effects and dummies for missing regressor observations.

Predictors of Feminine Conformity

Table A.2: Predictors of Feminine Conformity

	Girls		Boys	
	Univariate Model	Saturated Model	Univariate Model	Saturated Model
Cognitive Skills	-0.099***	-0.068***	-0.083***	-0.081***
	(0.016)	(0.019)	(0.011)	(0.012)
Externalising Behaviour	0.081***	0.028	0.043***	0.017
	(0.021)	(0.024)	(0.011)	(0.013)
Internalising Behaviour	0.066***	0.027	0.028***	-0.001
	(0.017)	(0.019)	(0.011)	(0.011)
Log Parental Income	-0.075**	-0.043	0.000	0.021
	(0.035)	(0.036)	(0.025)	(0.025)
Father's Age	-0.008***	-0.004	-0.003	-0.009***
	(0.003)	(0.004)	(0.002)	(0.003)
Mother's Age	-0.009***	-0.006	0.000	0.008**
	(0.003)	(0.005)	(0.002)	(0.003)
Birth Order	0.009	0.014	0.026**	0.016
	(0.016)	(0.019)	(0.011)	(0.014)
Mum's Hours Worked	-0.007	-0.049*	0.005	0.019
	(0.016)	(0.028)	(0.012)	(0.021)
Mum Employed	0.018	0.100*	-0.006	-0.033
	(0.032)	(0.058)	(0.023)	(0.042)
Female Headteacher	0.001	0.017	0.052	0.041
	(0.039)	(0.039)	(0.032)	(0.030)
Women: Employment	-0.027	-0.027	-0.017	-0.013
	(0.018)	(0.017)	(0.013)	(0.012)
Women: Went to Uni	-0.068***	-0.066***	-0.010	-0.010
	(0.018)	(0.018)	(0.013)	(0.013)
Dad: Post Compulsory	-0.138***	-0.055	-0.018	0.026
	(0.043)	(0.045)	(0.031)	(0.034)
Dad: Some College	-0.095	0.038	-0.014	0.034
	(0.080)	(0.091)	(0.059)	(0.071)
Mum: Post Compulsory	-0.120***	-0.037	-0.032	0.024
	(0.039)	(0.042)	(0.027)	(0.028)
Mum: Some College	-0.130	-0.051	0.032	0.082
	(0.104)	(0.100)	(0.076)	(0.088)
Expect to go to Uni	-0.086	0.001	-0.182**	-0.122
	(0.096)	(0.086)	(0.078)	(0.082)
Only Child	-0.066	0.017	-0.009	0.061
	(0.070)	(0.074)	(0.053)	(0.058)
Sisters Only	0.078*	0.115**	0.073**	0.124***
	(0.045)	(0.053)	(0.031)	(0.039)
Sisters & Brothers	0.036	0.043	0.008	0.031
	(0.037)	(0.047)	(0.027)	(0.035)

Notes: Sample size is 4,056 girls and 4,030 boys. We report significance at the 1% (***), 5% (**), and 10% (*) levels with robust standard errors. Columns 2 and 4 present coefficients from the regression of our feminine conformity index on all the listed determinants (the models in Figure 4). Columns 1 and 3 contain coefficients from a univariate regression of the index on the given determinant. We include dummies for missing regressors.

Summary Statistics for Determinants of Feminine Conformity

Table A.3: Summary Statistics for Determinants

	Mean	SD
Cognitive skills 11	0.05	1.00
Externalising Behaviour	-0.08	0.75
Internalising Behaviour	-0.04	0.92
Log Parental Income	3.71	0.55
Father's Ed Level	0.29	0.55
Dad: Age at birth	30.74	6.30
Mother's Ed Level	0.28	0.50
Mum: Age at birth	27.62	5.71
Only child	0.08	0.27
Sisters Only	0.21	0.41
Sisters and Brothers	0.45	0.50
Birth order	2.12	1.09
Aspiration for child: uni	1.78	0.48
Mum's hours worked	27.59	13.30
Mum employed	0.58	0.49
Female head teacher	0.21	0.41
Region: Female employment	14.24	1.16
Region: Female/male uni share	0.60	0.08

Notes: This table shows summary statistics for the determinants of feminine conformity used in Figure 4.

Relative Importance of Predictors

Table A.4: Girls – Shapley Values and Feature Importance

Determinant	Shapley Value	Shapley Rank
Cognitive Skills	0.0051	1
Women: Went to Uni	0.0033	2
Externalising Behaviour	0.0015	3
Internalising Behaviour	0.0014	4
Father's Education	0.0013	5
Mother's Age	0.0012	6
Father's Age	0.0011	7
Mother's Education	0.0010	8
Women: Employment	0.0007	9
Parent Exp: Uni	0.0007	10
Log Parental Income	0.0005	11
Mum's Hours Worked	0.0004	12
Sibling Composition	0.0004	13
Mum Employed	0.0003	14
Female Headteacher	0.0003	15
Birth Order	0.0002	16

Notes: The sample is 4,056 girls. Column 1 reports Shapley values, decomposing the share of the R^2 attributable to each variable in the saturated regression in Table A.2. Column 2 ranks determinants based on their Shapley value - a lower rank indicates a larger contribution to R^2. We include dummies for missings, whose contribution to the R^2 is held constant (and thus not computed) by including them across model combinations.

Relative Importance of Predictors

Table A.5: Boys – Shapley Values and Feature Importance

Determinant	Shapley Value	Shapley Rank
Cognitive Skills	0.0086	1
Father's Education	0.0024	2
Father's Age	0.0013	3
Externalising Behaviour	0.0012	4
Sibling Composition	0.0010	5
Mother's Age	0.0007	6
Birth Order	0.0006	7
Female Headteacher	0.0006	8
Women: Employment	0.0003	9
Internalising Behaviour	0.0003	10
Parent Exp: Uni	0.0003	11
Mother's Education	0.0002	12
Mum's Hours Worked	0.0002	13
Women: Went to Uni	0.0001	14
Log Parental Income	0.0001	15
Mum Employed	0.0001	16

Notes: The sample is 4,030 boys. Column 1 reports Shapley values, decomposing the share of the R^2 attributable to each variable in the saturated regression in Table A.2. Column 2 ranks determinants based on their Shapley value - a lower rank indicates a larger contribution to R^2. We include dummies for missings, whose contribution to the R^2 is held constant (and thus not computed) by including them across model combinations.

A.2 Tables for Boys

Life Cycle Trends in Earnings (Boys)

Table A.6: Life Cycle Trends in Earnings (Boys)

	Age 23	Age 33	Age 42	Age 50	Age 55
Panel A: Log Earnings in Survey Years					
Feminine Conformity Index	0.015	-0.007	0.009	-0.001	-0.040
	(0.009)	(0.016)	(0.019)	(0.022)	(0.029)
Panel B: Log Average Hourly Wages in Survey Years					
Feminine Conformity Index	0.004	-0.003	0.034*	-0.002	-0.032
	(0.009)	(0.016)	(0.019)	(0.022)	(0.025)
Panel C: Log Annual Hours Worked in Survey Years					
Feminine Conformity Index	0.005*	0.001	-0.005**	-0.006	0.003
	(0.003)	(0.002)	(0.002)	(0.004)	(0.006)
Panel D: Employment Status in Survey Years					
Feminine Conformity Index	0.001	-0.004	-0.003	-0.017**	-0.005
	(0.007)	(0.007)	(0.007)	(0.009)	(0.011)
Cognitive & Non-Cognitive Skills	✓	✓	✓	✓	✓
Parental Education & Income	✓	✓	✓	✓	✓
Family Background	✓	✓	✓	✓	✓
County Fixed Effects	✓	✓	✓	✓	✓

Notes: Sample size is 4,030 boys. 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.

Domain-Specific Feminine Conformity Indexes (FCIs) (Boys)

Table A.7: Domain-Specific Feminine Conformity Indexes (FCIs) for Boys

	Log Lifetime Earnings (1)	Share of Time Employed (2)	Log Lifetime Average Wages (3)	Log Lifetime Hours Worked (4)
FCI - Education	0.041	0.017**	0.070*	0.027
	(0.041)	(0.009)	(0.037)	(0.018)
FCI - Work	-0.030	-0.009**	-0.008	-0.025***
	(0.020)	(0.004)	(0.018)	(0.010)
FCI - Home Life	0.016	0.004	0.024	-0.003
	(0.024)	(0.005)	(0.022)	(0.010)
FCI - Hobbies	0.035	0.006	0.015	0.019**
	(0.022)	(0.005)	(0.019)	(0.010)
Cognitive & Non-Cognitive Skills	✓	✓	✓	✓
Parental Education & Income	✓	✓	✓	✓
Family Background	✓	✓	✓	✓
County Fixed Effects	✓	✓	✓	✓

Notes: Sample size is 4,030 boys. We report significance at the 1% (***), 5% (**), and 10% (*) levels with robust standard errors. Columns (3) and (4) are estimated on a sub-sample of 3,998 boys who are recorded as working in at least one month between ages 23 and 55, so wages and hours-worked are non-zero in levels. Each regression includes: indices for cognitive and non-cognitive skills and our full set of family background and geographic controls, including parental income and education, number of and sex composition of siblings, birth order, mother's and father's age at birth, parental aspiration, and family stability. Additionally, we control for essay length and the share of each essay spent on each of the four domains. We include 111 county fixed effects.

Share of Association Mediated through Education, Family and Occupational Choice

Table A.8: Share of Association Mediated for Boys

	Coefficient	
	Baseline	With Mediators
Log Lifetime Earnings	-0.007	-0.002
Log Avg Wage	-0.000	0.006
Log Total Hours Worked	-0.006	-0.003
Employment	-0.004	-0.002

Notes: Sample size is 4,030 boys. This table provides the analysis in the first two columns of Table 8 for the boys. We do include the 'Percent Explained by Mediator' columns, as there are no significant effects to explain.

Table A.9: Alternative Explanations: boys

	Lifetime Earnings	Lifetime Average Wages	Employed	Log Total Hours Worked
Panel A: Other Text Factors				
Feminine Conformity Index	-0.005	0.001	-0.004	-0.006
	(0.019)	(0.017)	(0.004)	(0.008)
Education Score	0.007	0.007	-0.000	0.001
	(0.019)	(0.016)	(0.004)	(0.008)
Cultivation Score	0.012	0.031*	-0.005	-0.015
	(0.020)	(0.018)	(0.005)	(0.012)
Affluence Score	0.002	0.001	0.000	0.001
	(0.019)	(0.018)	(0.004)	(0.010)
Status Score	0.000	0.009	-0.003	-0.011
	(0.017)	(0.014)	(0.004)	(0.008)
Morality Score	0.001	-0.011	-0.003	0.004
	(0.016)	(0.015)	(0.004)	(0.008)
Employment Score	-0.023	-0.013	-0.009**	-0.017**
	(0.015)	(0.014)	(0.004)	(0.007)
Panel B: Complementarities Between Covariates – Post-LASSO Model				
Feminine Conformity Index	-0.009	-0.002	-0.005	-0.008
	(0.018)	(0.016)	(0.004)	(0.009)
Panel C: Removing Negation Phrases				
Feminine Conformity Index	-0.004	-0.004	-0.003	-0.004
	(0.018)	(0.017)	(0.004)	(0.007)
Cognitive & Non-Cognitive Skills	✓	✓	✓	✓
Parental Education & Income	✓	✓	✓	✓
Family Background	✓	✓	✓	✓
County Fixed Effects	✓	✓	✓	✓

Notes: Sample size is 4,030 boys. 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 coefficient from a post-LASSO regression – we only report the index, which is always selected by LASSO (unlike non-cognitive skills). We add dummies for missing regressor observations and fixed effects to the post-LASSO specification.

A.3 Additional Figures

Figure A.1: Correlation of Feminine Conformity with Child's Preferred Activities at Age 11

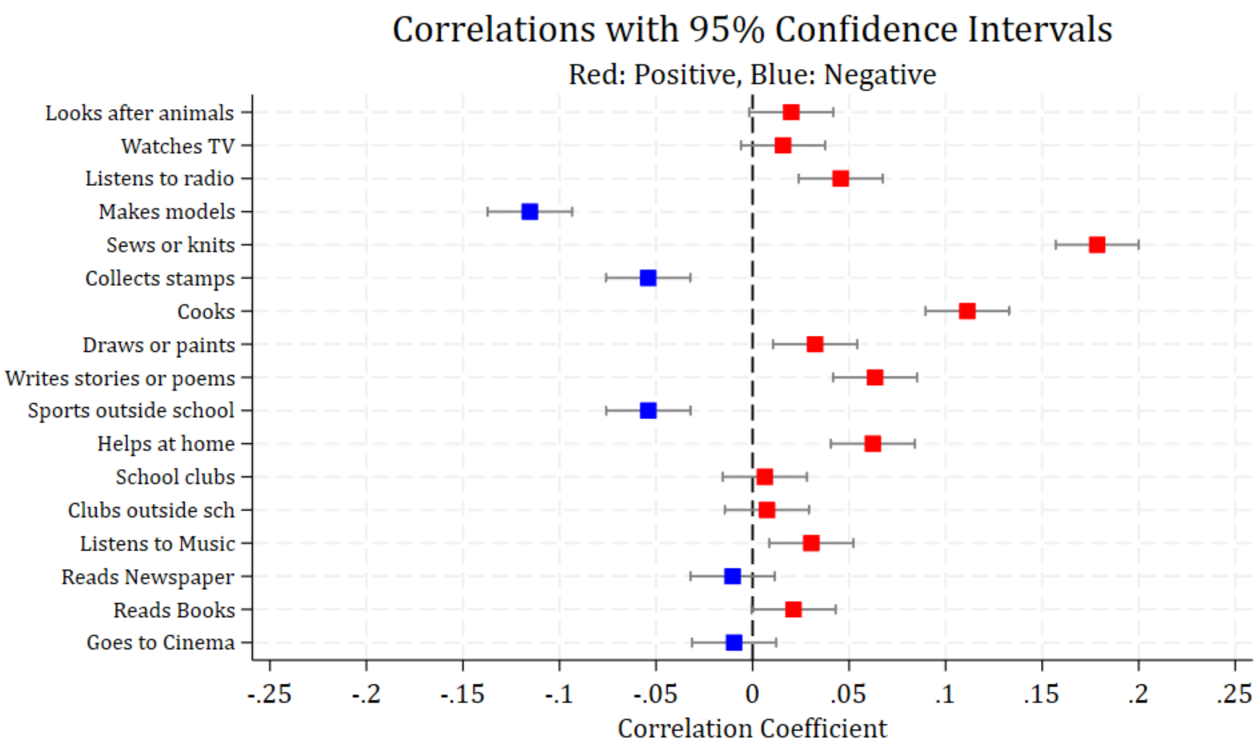
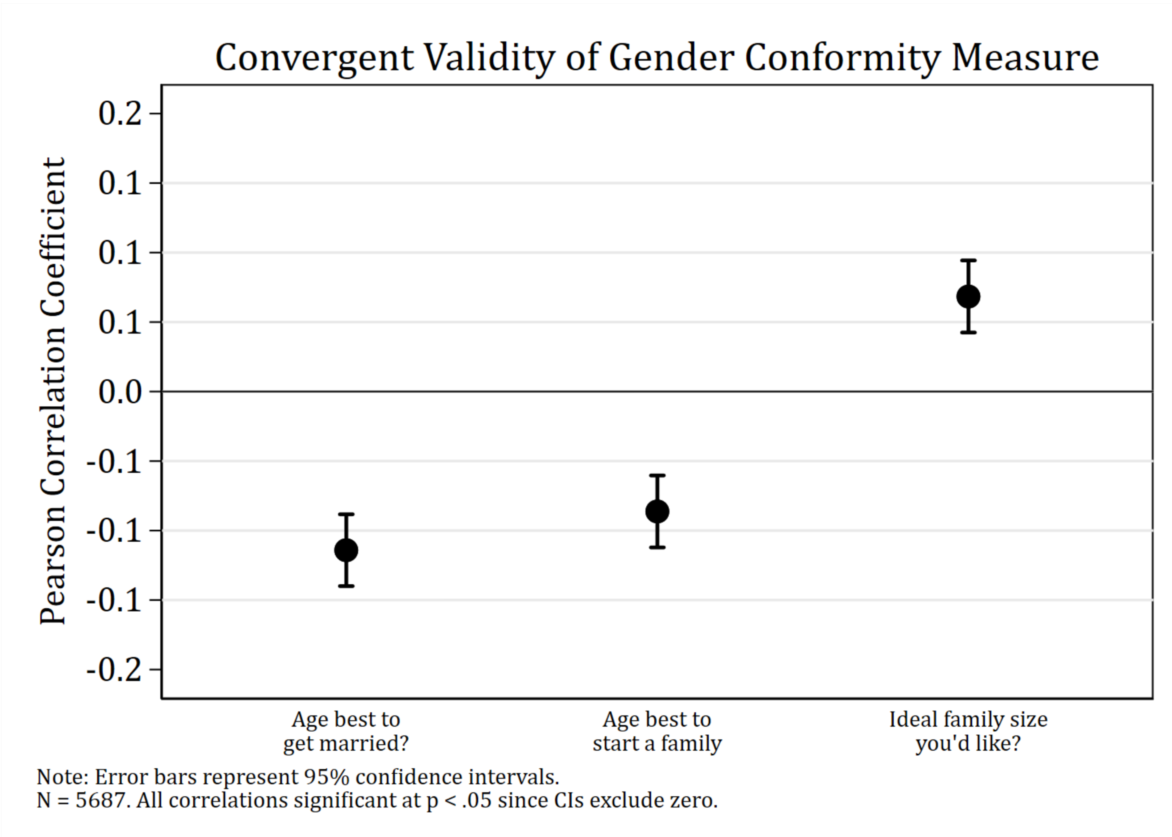


Figure A.2: Correlation of Feminine Conformity Index and Age 16 Measures of Attitudes Towards Family Formation



A.4 Robustness Checks

Additional Controls

Table A.10: Additional Controls

	Lifetime Earnings	Lifetime Average Wages	Employed	Average Weekly Hours Worked
Panel A: More Skills Controls				
Feminine Conformity Index	-0.038**	-0.022*	-0.004	-0.018*
	(0.017)	(0.011)	(0.004)	(0.010)
Panel B: School Quality Controls				
Feminine Conformity Index	-0.041**	-0.024**	-0.004	-0.018*
	(0.017)	(0.011)	(0.004)	(0.010)
Panel C: Aspired Occupation				
Feminine Conformity Index	-0.041**	-0.022*	-0.002	-0.015
	(0.017)	(0.011)	(0.004)	(0.010)
Panel D: Region Fixed Effects				
Feminine Conformity Index	-0.038**	-0.021*	-0.002	-0.016*
	(0.017)	(0.011)	(0.004)	(0.010)
Cognitive & Non-Cognitive Skills	✓	✓	✓	✓
Parental Education & Income	✓	✓	✓	✓
Family Background	✓	✓	✓	✓
County Fixed Effects	✓	✓	✓	✓

Notes: Sample size is 4,056 girls. We report significance at the 1% (***), 5% (**), and 10% (*) levels with robust standard errors. Columns (2) and (4) are estimated on a sub-sample of 3,998 boys who are recorded as working in at least one month between ages 23 and 55, so wages and hours-worked are non-zero in levels. 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 indices, 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).

Alternative Word Embedding Models

Table A.11: Alternative Word Embedding Models

	Trained WEM 1955-1959 (1)	Trained WEM 1960-1964 (2)	Trained WEM 1965-1969 (3)	Trained WEM 1970-1974 (4)	Trained WEM 1975-1979 (5)
Panel A: Girls					
Feminine Conformity Index	-0.043**	-0.052***	-0.040**	-0.033**	-0.027
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Panel B: Boys					
Feminine Conformity Index	0.005	0.013	0.022	0.017	0.013
	(0.018)	(0.018)	(0.019)	(0.018)	(0.018)
Cognitive & Non-Cognitive Skills	✓	✓	✓	✓	✓
Parental Education & Income	✓	✓	✓	✓	✓
Family Background	✓	✓	✓	✓	✓
County Fixed Effects	✓	✓	✓	✓	✓

Notes: Sample size is 4,056 girls and 4,030 boys. We report significance at the 1% (***), 5% (**), and 10% (*) levels with robust standard errors. In this table, we re-estimate the specification in Column (3) of Table 3 using alternative word-embedding models to construct the feminine conformity index. We train five alternative WEMs, trained on texts from non-overlapping five-year windows from 1955 to 1979 – 1955-1959, 1960-1964, 1965-1969, 1970-1974 and 1975-1979. 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, family stability and county fixed effects.

Index Construction

Table A.12: Index Construction

	Lifetime Earnings	Lifetime Average Wages	Employed	Average Weekly Hours Worked
Panel A: Occur $\geq$ 200 times, $\geq$ 2SD from neutral				
Feminine Conformity Index	-0.029	-0.026**	-0.004	-0.017
	(0.018)	(0.012)	(0.005)	(0.011)
Panel B: Occur $\geq$ 200 times, include all neutral words				
Feminine Conformity Index	-0.041**	-0.033***	-0.007	-0.026**
	(0.019)	(0.013)	(0.005)	(0.011)
Panel C: Occur $\geq$ 10 times, $\geq$ 1SD from neutral				
Feminine Conformity Index	-0.029	-0.026**	-0.004	-0.015
	(0.020)	(0.013)	(0.005)	(0.012)
Panel D: Occur $\geq$ 100 times, $\geq$ 1SD from neutral				
Feminine Conformity Index	-0.036*	-0.028**	-0.006	-0.020*
	(0.019)	(0.013)	(0.005)	(0.011)
Panel E: Occur $\geq$ 300 times, $\geq$ 1SD from neutral				
Feminine Conformity Index	-0.040**	-0.027**	-0.006	-0.024**
	(0.019)	(0.012)	(0.005)	(0.011)
Cognitive & Non-Cognitive Skills	✓	✓	✓	✓
Parental Education & Income	✓	✓	✓	✓
Family Background	✓	✓	✓	✓
County Fixed Effects	✓	✓	✓	✓

Notes: Sample size is 4,056 girls. 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 1 standard deviation 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 feminine conformity index. For our baseline results, 285 different words are used to construct the feminine 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.

Alternative Measures of Lifetime Earnings

Table A.13: Alternative Measures of Lifetime Earnings

	Lifetime Earnings (1)	Log Lifetime Earnings of Employed (2)	Log Lifetime Observed Earnings (3)	Log of Simple Average of Lifetime Earnings (4)	Log Simple Average of Observed Lifetime Earnings (5)
Panel A: Girls					
Feminine Conformity Index	-14327.091***	-0.034**	-0.049**	-0.032**	-0.031**
	(5382.162)	(0.017)	(0.024)	(0.013)	(0.012)
Cognitive Skills	79337.437***	0.187***	0.203***	0.125***	0.180***
	(6880.937)	(0.021)	(0.029)	(0.017)	(0.015)
Externalising Behaviour	9784.269	0.069**	0.091**	0.032	-0.005
	(7836.783)	(0.030)	(0.039)	(0.023)	(0.021)
Internalising Behaviour	-19019.339***	-0.067***	-0.005	-0.024	-0.028*
	(5953.853)	(0.023)	(0.031)	(0.018)	(0.016)
Panel B: Boys					
Feminine Conformity Index	6435.400	-0.005	-0.023	-0.003	0.005
	(16041.396)	(0.018)	(0.023)	(0.017)	(0.016)
Cognitive Skills	132478.004***	0.150***	0.150***	0.129***	0.133***
	(14254.523)	(0.015)	(0.020)	(0.014)	(0.012)
Externalising Behaviour	-20719.324*	-0.028*	-0.051**	-0.003	-0.013
	(12205.641)	(0.017)	(0.025)	(0.015)	(0.012)
Internalising Behaviour	-41707.992***	-0.052***	-0.064***	-0.039***	-0.047***
	(10854.087)	(0.015)	(0.022)	(0.013)	(0.010)
Parental Education & Income	✓	✓	✓	✓	✓
Family Background	✓	✓	✓	✓	✓
County Fixed Effects	✓	✓	✓	✓	✓

Notes: Sample size is 4,056 girls and 4,030 boys. We report significance at the 1% (***), 5% (**), and 10% (*) levels with robust standard errors. In this table, we re-estimate the specification in Column 3 of Table 3 using alternative measures of lifetime earnings. Column 1 estimates the regression with earnings in levels, using our main sample, winsorising at the 1% and 99% levels. Mean lifetime earnings are £384,682, so the implied elasticity for girls from column 1 is -14,327/384,682=-0.037. Column 2 conditions on being employed at least once between age 23 and age 55, and therefore drops those with zero lifetime earnings - the sample size is 3,966 girls and 3,998 boys. Columns 3-5 test alternative imputation methods for lifetime earnings. Column 3 uses a measure of lifetime earnings identical to our main measure, except that it only uses earnings that are observed in the NCDS data (that is, without imputing for missing earnings based on individual fixed effects and educational attainment interacted with time fixed effects). Column 4 takes the log of a simple mean across earnings, but uses both observed earnings and imputed ones. Column 5 takes the simple mean across earnings observed at each survey wave (if earnings are unobserved, then that survey wave is excluded from the average). Thus columns 4 and 5 do not use the employment histories and account for periods of non-employment between survey waves. 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 include 111 county fixed effects and dummies for missings.

B Data

B.1 Attrition and 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 1958. 15,335 of these individuals are surveyed at age 11, of which 13,675 cohort members wrote essays. 10,551 of those essays were successfully digitized. We impose two restrictions on those remaining in the sample. First, we drop 1,129 individuals who were not surveyed at ages 23, 33, 42, 50 or 55. Finally, 1,296 individuals are excluded because we never observe their employment status. Thus, we arrive at our final analysis sample of 8,086 individuals, of which 4,056 are girls.

Table B.1: Sample Selection Criteria

Sample Selection Criteria	# of Individuals
Total Sample in 1958	18,558
Surveyed at Age 11	15,336
Wrote Essays at Age 11	13,675
Successfully Digitized Essays	10,511
Observed at least once in Adulthood	9,382
Employment status available at least once (ages 23-55)	8,086
Analysis Sample	**8,086**
Girls	4,056
Boys	4,030

Notes: Here, we present our sample selection criteria, and the number of individuals who remain after cumulatively applying our criteria. For instance, 13,675 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.

To study the potential implications of attrition in the NCDS data, we split the total sample from 1958 of 18,558 individuals into three groups: those not surveyed at age 11, those observed (i.e., surveyed) at age 11 but not in our analysis sample, and our analysis sample.

Panel A of Table B.2 contains balance tests for these three groups. We include variables observed at birth that have been shown to be determinants of later life outcomes: parental education, birth weight, and parental socioeconomic status. In Panel B, we present additional balance tests for four age-11 variables: cognitive skills, externalising behaviour, internalising behaviour and parental income, as well as the feminine conformity index and lifetime earnings.

Table B.2: Sample selection

	Unobserved at 11	Observed at 11	Analysis Sample
Panel A: Age 0 Variables			
Birthweight	3.620	3.681	3.642
	(2.492)	(2.024)	(1.872)
Mother: Post-Comp	0.244	0.218	0.236
	(0.430)	(0.413)	(0.425)
Mother: Uni	0.011	0.024	0.024
	(0.106)	(0.154)	(0.155)
Father: Post-comp	0.193	0.177	0.195
	(0.395)	(0.382)	(0.396)
Father: Uni	0.063	0.045	0.046
	(0.243)	(0.206)	(0.210)
Socio-Economic Status	4.170	3.854	4.001
	(4.206)	(3.927)	(4.003)
Panel B: Age 11+ Variables			
Cognitive Skills at 11		-0.110	0.080
		(1.112)	(1.029)
Internalising Behaviour at 11		-0.087	0.060
		(1.238)	(0.993)
Externalising Behaviour at 11		-0.094	0.069
		(1.174)	(0.902)
Log Parental Income at Age 16		3.673	3.718
		(0.586)	(0.556)
Feminine Conformity Index		-0.020	-0.002
		(0.902)	(0.936)
Log Lifetime Earnings		13.017	13.000
		(1.086)	(1.064)

Notes: This table compares characteristics of individuals in the NCDS data. Column 1 refers to the 3,222 individuals who were not surveyed, and thus did not write an essay, at age 11. The second column refers to 7,250 individuals who were observed at age 11, but who were disqualified from our final sample because they: (i) did not write an essay, (ii) were not surveyed in adulthood, or (iii) have unknown employment status between age 23 and 55. The final column consists of 8,086 individuals in our analysis sample. We report means and standard deviations of all variables. Panel A presents variables observed at age 0 (and thus are available for the full sample of 18,558), while Panel B presents variables recorded at (or after) age 11. Among age 0 variables, father's post compulsory education and family socio-economic status are statistically distinguishable between columns 2 and 3 at the 5% level. The differences in all three skills variables and parental income between columns 2 and 3 are statistically significant at the 1% level – individuals in our analysis sample have higher cognitive and non-cognitive skills, and parents with higher income.

Panel A of Table B.2 shows that our analysis sample is very comparable on baseline characteristics to excluded individuals who were surveyed at age 11. Relative to those not surveyed at age 11, our analysis sample has higher (lower) rates of mothers (fathers) attaining post-compulsory education, but is otherwise comparable. Panel B suggests that our analysis sample has slightly better cognitive and non-cognitive skills than excluded individuals who were surveyed at age 11, as well as slightly higher parental income. We control for all these variables in our analysis to account for this imbalance. Reassuringly though, our feminine conformity index and lifetime earnings are not statistically different between the two samples suggesting that our sample selection procedure is not inducing selection on our main variables.

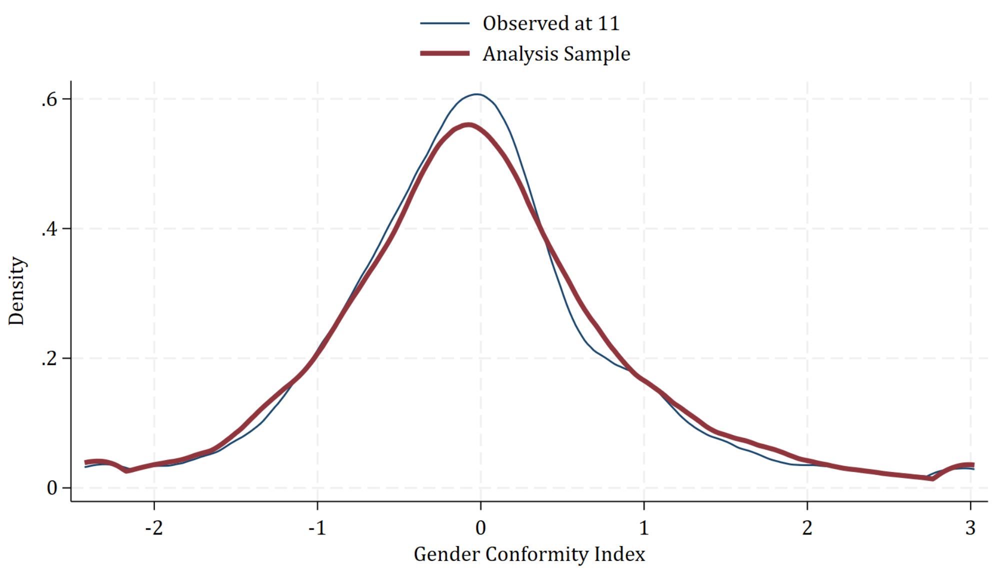


Figure B.3: Distribution of Feminine Conformity Index Across Samples

Next, we investigate whether attrition varies systematically with feminine conformity. Focusing first on attrition at age 11, Figure B.3 plots our index of feminine conformity for the 2,425 individuals who wrote essays at age 11 but were excluded from our sample, versus the 8,086 individuals in our analysis sample. There is no systematic difference in these distributions, and we find no significant difference (at the 1% level) in means, medians, or variances.

Finally, we investigate whether, conditional on being in the analysis sample, feminine conformity predicts the number of times the individual was interviewed in adulthood. Individuals could be interviewed up to five times during adulthood (at ages 23, 33, 42, 50, 55), but were often interviewed fewer times due to survey non-response. Table B.3 documents the relationship between this non-response in adulthood and our index of feminine conformity. Row (1) reports the average number of surveys responded to in adulthood for those with above versus below average levels of feminine conformity. These averages are statistically indistinguishable.

We impute earnings and wages based on item or survey non-response. Rows (2) and (3) of Table B.3 document the average number of non-imputed observations for earnings and wages per individual (with a maximum of 5). This can be smaller than the number of times the individual was in the survey due to item non-response. We find statistically significant differences between the groups. In particular, there is a positive association between our feminine conformity index and missing earnings and wages. Columns (3)-(5) of Table A.13 show robustness of our lifetime earnings result to any imputation procedure.

Table B.3: Non-response in Adulthood

	Below Average Feminine Conformity	Above Average Feminine Conformity
Average Number of Surveys	4.106	4.087
	(1.126)	(1.140)
Number of Non-Imputed Earnings Observations	3.233	2.980
	(1.446)	(1.504)
Number of Non-Imputed Wage Observations	3.156	2.914
	(1.445)	(1.501)

Notes: This table compares non-response in adulthood and imputation for the 8,086 individuals in our analysis sample. The sample is split in half based on median feminine conformity. We report means and standard deviations of all variables. Row (1) reports the average number of successful surveys conducted (per individual) in adulthood. The difference is statistically insignificant, even at the 1% level (p-value = 0.43). Rows (2) and (3) report the average number of non-imputed observations for earnings and wages per individual (maximum of 5). The difference across both sub-samples is statistically significant at the 1% level.

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 (Cunha and Heckman, 2008; Agostinelli and Wiswall, 2016), 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 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, minimizing measurement error in the final score. The indices of cognitive skills, internalizing behavior and externalizing behavior we use are these Bartlett scores.

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.

For internalizing and externalizing behavior, we use the measures proposed by Papageorge et al. (2024), originally suggested in Ghodsian (1977). A key distinction is that we weight each measure by its noise when aggregating, while Papageorge et al. (2024) use equal weights for all measures. For externalizing behavior, we use measures which record whether the cohort member is unconcentrated, fidgety, miserable, or destructive, which correlate negatively with the index, as well as measures of perceived obedience and socializing skill which correlate positively. For internalizing behavior we use measures which record whether a child is anxious for acceptance or is hostile towards adults or children, writes off adults and acts without regard for others. These measured behaviors are as defined by the Bristol Social Adjustment Guide.

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.[27] 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 logarithm of the sum of the three net, predicted earnings variables in the NCDS is our measure of parental income. For parents who did not answer any of the earnings questions, parental income is set as missing, following Blanden et al. (2007), who treat this similarly.

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 and earnings at the time of a survey. We model wages and earnings as a function of an individual-specific fixed

[27]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

effect, a time-fixed effect and an interaction between years of schooling and time:

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

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 or wages 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). Earnings or wages recorded as zero when the individual is recorded as in work are coded as missings, since they are attributed to item non-response and may be imputed. We set (imputed) values to zero whenever $\widehat{Y_{it}} < 0$. Monthly earnings (wages) are calculated from earnings (wages) at each survey date by imputing a linear trajectory between each survey data point for every month between surveys. These are adjusted using the monthly Activity History data; months spent out of work are assigned zero earnings and missing wages. The sum of monthly earnings is our estimate of lifetime earnings and the mean of monthly wages estimates average wages over the lifecycle.

B.5 Quality of Life and Health

Below, we elaborate on our age 55 measures of quality of life, mental health and physical health.

Quality of Life We use Hyde et al. (2003)'s Control, Autonomy, Self-realization and Pleasure (CASP) scale to measure quality of life (QoL). The scale is designed for older adults, grounded in needs satisfaction theory and sociological concepts of successful aging. We use the CASP-12 score which consists of twelve 4-point Likert-scaled items spanning four theoretical domains: control, autonomy, self-realization, and pleasure. This version demonstrated superior empirical performance in capturing QoL among older adults in England and Wales (Wiggins et al., 2008). The CASP-12 is widely implemented in major longitudinal aging studies, including the Survey of Health, Ageing and Retirement in Europe and the English Longitudinal Study of Ageing.

Mental Health We measure mental health using the Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS). The scale comprises of 14 positively-worded statements covering aspects of mental health such as: experiencing pleasant emotions, satisfying interpersonal relationships, and the capacity for healthy psychological and behavioral growth. Respondents rate each statement on a 5-point Likert scale based on their experience over the previous two weeks. The ratings are summed across the 14 statements; we then standardize this sum to have mean zero and variance one. WEMWBS has robust psychometric properties, including high internal consistency and test-retest reliability (Tennant et al., 2007; Stewart-Brown et al., 2009). The scale has been extensively implemented in population health monitoring and intervention studies across the UK and internationally (Ng Fat et al., 2017; Clarke et al., 2011). It is recommended by the UK's National Health Service for assessing mental wellbeing in public health contexts.

Physical Health We measure physical health using a score derived from the RAND 36-Item Health Survey (also known as the SF-36), a widely used generic health status instrument (Hays et al., 1993). The score consists of ten items assessing limitations in performing various physical activities due to health problems, ranging from vigorous activities such as running to basic tasks such as bathing and dressing. Response options capture the degree of limitation experienced (limited a lot, limited a little, or not limited at all). Each item response is assigned a predetermined weight based on the original scaling algorithm, and responses are summed to create a raw score (Ware Jr, 2000). We standardize this sum to create a mean-zero index with unit variance. The physical functioning sub-scale of the SF-36 has demonstrated excellent reliability and validity across diverse populations (McHorney et al., 1993; Brazier et al., 1992), especially used to evaluate functional health among older adults (Walters et al., 2001).

C Estimating Feminine 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 feminine 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, where sentences are 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 C.4. When making these corrections, each student can infer the word's use in the context of the sentence.

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 using Python's *nltk* library. First, all stop words e.g., "the", "a", "an", "in", "and", "is"), which are pre-defined in the library, are removed.[28] The essay text is then vectorized into a matrix of 1-grams. This is our bag-of-words matrix. To this, we append counts of 2-grams that start with "not" or "no". 1-grams and 2-grams which appear less than 200 times across all essays are then removed. Thus, we arrive at our final representation for the essays.

Of note is our manual approach to correcting misspelled words. Whilst automatic spelling correction algorithms are available, they are ill-suited to the phonetic nature of children's

[28] Negation words, such as "not" and "no", are specifically excluded from the list of stop words.

Figure C.4: 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.

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, many standard spell-check software would not correctly replace these misspelled words.

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

C.2.1 Selecting Texts

The data we use to train our skip-gram Word2Vec algorithm is the **Google Books N-grams Version 2**; specifically, we use the British English 5-grams data.[29] This comprises of all 5-grams extracted from over 140,000 digitized texts written between 1958 and 1978. Below, we motivate our choice of training corpus, discussing its known limitations. We highlight that our

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

results are robust to time windows before or around when the essays were written, but are less robust to off-the-shelf alternatives trained on contemporary text corpora.

Text Corpus The Google Books N-grams dataset is derived from Google's book digitization project, which has scanned over 8 million books representing approximately 6% of all books ever published (Lin et al. (2012)). For our analysis, we use the 5-grams from all texts published between 1958 and 1978. This encompasses over 140,000 texts and primarily includes published books from university libraries and newspapers, but excludes magazines, pamphlets, and unpublished materials.[30] Across these texts, Google extracts sequences of n consecutive words (n-grams) along with their frequency of occurrence. The Version 2 dataset, which we use, contains all n-grams that appear at least 40 times across the corpus, helping to filter out optical character recognition (OCR) errors and extremely rare words and phrases.

The main advantage of these data is that their scale provides the statistical power necessary to learn robust word representations. WEMs require millions of word co-occurrences to accurately estimate the 300-dimensional vectors we assign to each word. The corpus is large enough that our 1958-1978 subset contains sufficient data to achieve this precision, while still zooming in on period-specific word meanings and cultural associations (Hamilton et al., 2016; Kozlowski et al., 2019). An important limitation is that any historical text corpus reflects publishing biases of its era. In mid-20th century Britain, published authors were disproportionately educated, white, and male, meaning our WEM captures gender norms as expressed in this published discourse (Michel et al., 2011; Pechenick et al., 2015). Moreover, since the corpus captures formal, edited language, which rarely features much colloquial language, slang from the essays may imperfectly characterized.

We choose to use 5-grams from the Google N-grams data to train our WEM. This is because the skip-gram algorithm we use for training (described below) learns word meanings by predicting context – longer phrases provide richer information about word co-occurrence patterns within context windows, improving the model's ability to capture semantic relationships

[30] The entire corpus consists of 8 million texts, of which 1.6 million are British English; of those, over 140,000 are published between 1958 and 1978.

(Mikolov et al., 2013). Although we train the skip-gram algorithm on data containing 5-grams, our model ultimately assigns vector representations only to individual words (1-grams), which we subsequently use to construct our feminine conformity index.

Time-Period We train our WEM on texts published between 1958 and 1978. In Table A.11, we show that our results are robust to using WEMs trained on texts from various time windows around when the essays were written (in 1969). Our results remain qualitatively the same (but become attenuated and lose statistical significance) when we use text corpora more than a decade past when the essays were written (e.g. from 1975-79 or 1980-89). When we use Google's Word2Vec or Stanford's GloVe WEMs, which are trained on extensive corpora of predominantly modern, digitized texts, our coefficient on the feminine conformity index in our lifetime earnings regression is even closer to zero (albeit, still negative). This is consistent with the findings in Kozlowski et al. (2019), who show that gender connotations in English have evolved throughout the twentieth century. For example, occupational terms like "nurse" have strong female associations in our WEM trained on texts from 1958 to 1978. However, these weaken in WEM trained on texts between 1980-89 (cosine similarity halves), and are close to gender neutral in Google's Word2Vec WEM or the GloVe WEM. This changing meaning of words leads to measurement error if the WEM is trained on data from a different period than when the essays were written, leading to attenuated parameter estimates.

C.2.2 Procedure and Specifications

Skip-Gram Algorithm Our word-embedding model (WEM) is a by-product of training Google's Word2Vec skip-gram algorithm on the previously described text corpus. The purpose of this skip-gram algorithm 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 dress". To make such predictions, this algorithm performs unsupervised learning. In other words, it "learns" about text patterns from data used to train it (in our case, the Google Books N-grams data), 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 algorithm assigns each word a numerical vector value; words which predict similar adjacent words are assigned similar vector values. A WEM is thus mapping from the words to the real-valued vector. We choose the Word2Vec skip-gram algorithm because it has been shown to perform best when trained on large text corpora which contain words that are used infrequently, making it well-suited to the substantial but varied Google Books dataset (Sabra and Sabeeh, 2020).

Our Specification 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 48-core cluster and takes around 76 hours to train.

C.3 Step 4: Constructing Gender Dimension

C.3.1 Antonym Pairs

Our gender dimension is constructed by differencing the embedding vectors of pairs of gender antonyms (e.g. “female” minus “male”, “her” minus “him” etc.). A potential alternative could be to construct two vectors – one averaging across “feminine” words, and the other averaging across “masculine” words. Below, we motivate our choice of using antonym pairs.

First, by constructing antonym pairs, we follow established practices in the word embedding literature. Word embedding models (WEMs) have been shown to solve analogy problems: e.g., “man is to woman as king is to —” achieving accuracy rates of 70-90% (i.e., correctly predicting ”queen”) when using the antonym pair approach (Mikolov et al. (2013)). This method has become standard in social science applications of WEMs. We follow Kozlowski et al. (2019), an influential sociology paper that justifies using antonym pairs by citing evidence from Mikolov et al. (2013) and related work. We use the same gender dimension vector as Kozlowski et al.

(2019). We view it as a strength of our analysis that we do not exercise discretion in constructing our own gender dimension vector.

Second, the issue with the alternative of a separate feminine and masculine dimension vector is that words like "man" and "woman" appear in texts in contexts that are not inherently gendered, but simply relate to people. Consequently, embedding vectors for "man" and "woman," (or "male" and "female" or "he" and "she") capture not only gender-specific associations but also person-specific associations. Without differencing, we conflate the gender dimension with a human dimension. By contrast, differencing antonym pairs isolates the component capturing feminine (versus masculine) associations, because shared human-related associations (e.g., with "person", "head", "alive") are eliminated via differencing.

We demonstrate this second point empirically. We construct two separate dimension vectors: one using only masculine words from our gender dimension vector (e.g., "man," "he") and another using only feminine words (e.g., "woman," "she"). We then create male and female indices using the same methodology as our feminine conformity index. Essays with language similar to the female vector receive high female index values, while those with language similar to the male vector receive high male index values.

These two indices are highly *positively* correlated (0.92), indicating that both capture human activities generally rather than distinctly gendered content. Moreover, each shows weaker correlation with our feminine conformity index: the female index correlates at 0.61, while the male index correlates at 0.37. This pattern of correlations confirms that separate dimension vectors fail to isolate gender-specific associations from human associations.

C.4 Validating our Procedure

C.4.1 Antonym Pairs Approach

We construct our gender dimension using antonym pairs. We implement Bolukbasi et al. (2016)'s approach of checking whether the gender dimension of our Word-Embedding Model, $\vec{GD}$, can identify pairs of gendered words such as $\vec{king} - \vec{queen}$. To do this we use select the fifteen word pairs whose differenced vector values lie closest to $\vec{GD}$ (as measured by their cosine similarity), first using all possible word pairs in our Word-Embedding Model and second to all word pairs in the essays. Figure C.5 illustrates our results. The fact that many of these pairs are explicitly gendered validates that $\vec{GD}$ indeed captures the feminine-masculine distinction.

Figure C.5: Pairs of Gendered Words

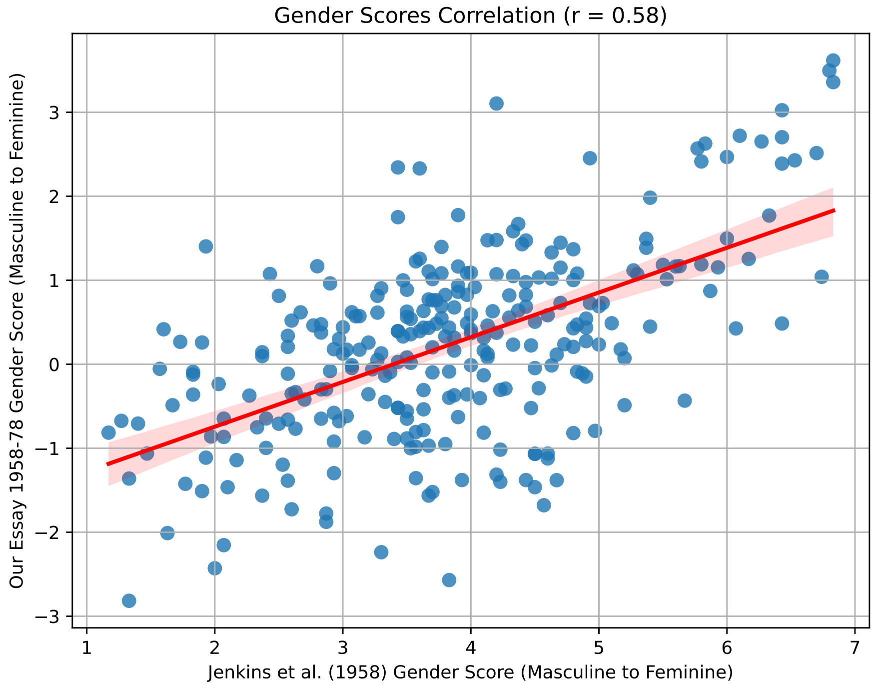
(a) All words ($\approx$ 2 million) (b) Words in essays ($\approx 18,000$)

Notes: Word clouds of the 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.

C.4.2 Gender Profiles of Words

Jenkins et al. (1958) ask 540 psychology students to rate the gender connotations of 360 words and phrases. Participants had to assign a number between 1 and 7 to each word, where a higher rating meant they considered the word more feminine. 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, Figure C.6 plots the rank correlation between the projection values our WEM assigns and the corresponding average rating for each word assigned by participants of Jenkins et al. (1958)'s study. Reassuringly, this turns out to be 0.58, 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.

Figure C.6: Correlation - WEM Projection Values and Jenkins Jenkins et al. (1958)'s Scores



C.4.3 Simpler Indices

To stress test our index of feminine conformity, we construct an alternative, simpler index based on (unweighted) word counts of a list of feminine words in our essays that we select independently of our WEM.

We selected the following words from the intersection of words that all four authors agreed have strong gender connotations: husband, house, boy, girl, little, baby, stay, cook, wash, clean, breakfast, knit, tidy, kitchen, housework, sew, hostess, wedding, and washing. 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 feminine conformity.

We find that this naive index has a correlation of 0.49 with our feminine conformity index for our pooled analysis sample. However, the naive index only explains 24% of the variation in our feminine conformity 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, we construct a weighted version of the naive index, using projection values from our word embedding model. Both the R^2 and raw correlation are largely unchanged (0.50 and 25%, respectively), 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. Moreover, words identified as gendered by our baseline procedure are more likely to reflect 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 (current) experiences of gender norms.[31]

C.4.4 Measurement Error in Feminine Conformity

Our procedure for estimating feminine conformity relies on a gender dimension vector, which averages 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).

While our averaging procedure may reduce measurement error, it may not eliminate it. To assess the role of measurement error, we construct ten different feminine conformity indices, each based on only one of our ten pairs of gendered words. We use these ten feminine conformity indices as instruments for our baseline index.[32] Appendix Table C.4 shows that when instrumenting, the coefficients on feminine conformity for our four main labor market outcomes remain largely unchanged. This affirms that measurement error in our gender dimension is unlikely to be of first-order importance.

C.4.5 Discussion

Words that are more "gendered" (either more masculine or feminine) or that appear more frequently in an essay will matter more for an essay's feminine conformity index value. We investigate the drivers of our feminine conformity index in two separate ways, as described

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

[32] This is analogous to Agostinelli and Wiswall (2025)'s approach of using one measure of cognitive skill as an instrument for another.

Table C.4: Instrumenting for Feminine Conformity

	Lifetime Earnings (1)	Lifetime Average Wages (2)	Employed (3)	Average Weekly Hours Worked (4)
Panel A: Girls				
Feminine Conformity Index	-0.039**	-0.022*	-0.004	-0.018*
	(0.018)	(0.012)	(0.004)	(0.010)
Panel B: Boys				
Feminine Conformity Index	-0.007	0.001	-0.006	-0.010
	(0.020)	(0.018)	(0.004)	(0.008)
Cognitive & Non-Cognitive Skills	✓	✓	✓	✓
Parental Education & Income	✓	✓	✓	✓
Family Background	✓	✓	✓	✓
County Fixed Effects	✓	✓	✓	✓

Notes: Sample size is 4,056 girls. We report significance at the 1% (***), 5% (**), and 10% (*) levels with robust standard errors. In this table, we instrument for the feminine conformity index using 10 noisy measures of feminine 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.

below.

LASSO Approach First, we use our word-embedding model to calculate the cosine similarity of every word with the gender dimension. Next, we multiply the projection value of each word (that is, the standardized cosine similarity) with how frequently it appears in each essay. We then estimate a LASSO regression of our index of feminine conformity on these weighted word counts. The twenty words that arise as the most important predictors of our index are: husband, female, time, girl, home, mother, day, wife, school, male, house, went, baby, team, job, hair, sister, tea, married and army. It is reassuring that many of these words might be considered gendered.

Words Driving High versus Low Index Values We explore an alternative approach to examining the words driving our feminine conformity index. We first restrict our attention to essays that appear in the top and bottom 25% of the feminine conformity index distribution. We then count the total number of times each word appears in essays that are in the top 25%

of the index distribution, and the total number of times each word appears in essays that are in the bottom 25%. Next, we weight counts of each word by its projection value from the WEM. Finally, we calculate the difference in weighted frequency (by word) between the bottom 25% and the top 25% of the index distribution. For instance, if "lady" has a projection value of 1.2 and appears 5 times across the top 25% of essays and twice across the bottom 25%, the difference in weighted frequency for "lady" would be $1.2(5-2) = 3.6$. The twenty words with the highest weighted frequency, and therefore whose presence explains why an essay is in the top versus bottom 25% of the index's distribution, are: female, husband, girl, time, baby, mother, team, day, hair, male, married, went, job, home, sister, little, dress, army, good, and nice.

Examples of Words Here, we present examples of categories of words in our essays and how they are classified by our gender dimension and word-embedding model.

1. **Place Names** - Some essays mention place names, including "London", "Manchester", "Paris", "Germany" etc. These consistently project toward the masculine end of our gender dimension vector. This is likely because they frequently appear in contexts associated with work, careers and professional ambition in the training corpus: domains that have historically been male-dominated.

2. **Occupations** - Several essays mention specific job titles, including "nurse", "teacher", "engineer", "policeman" etc. Some, such as nurse, hostess and teacher, have female associations. Most occupations though, even seemingly gender-neutral occupations like "secretary" or "manager", are attributed male associations, likely due to the period between 1958 and 1978 being characterized by relatively low levels of female employment.

3. **Errands** - Although not frequently mentioned, some essays mention locations associated with daily errands including "supermarkets", "pharmacy", "bakery" etc. Such locations range from being gender-neutral to having mild female associations, likely because they appear more frequently in the context of home production, which in our training corpus is female-dominated.

C.5 Domain-Specific Feminine Conformity Indices

Femininity is a complex concept that encompasses multiple domains. To better understand which domains drive our feminine conformity index and its relationship with earnings and other outcomes, we construct domain-specific feminine conformity indices. We proceed in two steps. First, we use an LLM to extract the four topics that are most written about across all 10,511 essays. To do so, we use the following prompt which is asked to the LLM and explained in the box below– "LLM Prompt - Which Topics?". The LLM proposed four main topics: education, work and employment ("work"), hobbies and social life ("hobbies"), and home life. The corresponding four sub-topics per main topic that were generated by the LLM are enumerated below.

LLM Prompt - Which Topics?

CONTEXT # #

Below, is a list of essays written by 11-year-old children in the UK in the year 1969, in response to the prompt

> "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."

INSTRUCTIONS# #

The essays will appear below with each individual essay enclosed in `<< >>`. After examining all essays, give the 4 topics that are most prevalent across the essays. Topics should be a word or phrase, but not a sentence. Give 4 main subtopics within the topics. For instance, if the topic is hobbies, list the 4 most prevalent hobbies.

ESSAYS # #

`<<{ESSAYS}>>`

- Education
  - Studying to become an engineer, architect or other technical occupation
  - Learning new skills and hobbies
  - Pursuing personal interests and passions at university
  - Reading and learning (e.g. reading textbooks, taking a course)

- Work and Employment
  - Engineer (e.g. civil engineer, mechanical engineer, mechanical engineer)
  - Working as a teacher, nurse or healthcare professional
  - Working in a shop, office or retail environment
  - Pursuing a career in the arts or entertainment
- Hobbies and Social Life
  - Sports (e.g., football, tennis, swimming, rugby)
  - Music and art (e.g., playing instruments, painting, drawing)
  - Friendships and social relationships (e.g. shopping with friends)
  - Gardening and outdoor activities (e.g. growing vegetables, walking the dog)
- Home Life
  - Marriage and having children
  - Family life and responsibilities
  - Having a family home or a place to live with loved ones
  - Travel and exploration (e.g., going on holidays with family)

We then ask the LLM to categorize each contiguous block of text in each essay into one of five categories, the four described above and a miscellaneous category ("other"). Categorization is not necessarily unique; a given phrase or sentence can be at most in two categories. All words from the essay are preserved in this categorization. To construct our indices, we only keep blocks of text assigned to a single category, and ignore blocks of text assigned to "other". Our exact prompt is given below - "LLM Prompt - Split into Topics".

LLM Prompt - Split into Topics

INSTRUCTIONS

Below, enclosed in `<< >>`, is an essay written by an 11-year-old child in the UK in the year 1969. The essay responds to the prompt:

> "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."

Split the essay into CONTIGUOUS text blocks.
A block ends ONLY when the THEME CATEGORY changes. Include every word from the ORIGINAL ESSAY (no additions, deletions, or paraphrasing).

CATEGORIES *(choose from these codes; up to two per block if it CLEARLY discusses both)*

- (i) work/employment
- (ii) education
- (iii) hobbies/social life
- (iv) home life
- (vi) other

RESPONSE RULES
Only split when the theme changes. Try to keep each block in one theme; use two themes only if the text clearly contains both at once. Categorize based on what information the text conveys about the writer's activities, routines, preferences, or plans. Do not omit quotes, punctuation, or parentheticals. Preserve capitalization and spelling.

For the nth text block, respond EXACTLY in the following format:

> n.
> A. [Text block]
> B. [Theme category or categories]
> C. N/A

ESSAY TO EXAMINE
`<<{ESSAY}>>`

C.6 Other Latent Dimensions

In Section 4.4 (Panel A in Table 9), 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, righteousodious, 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 feminine conformity index and outcomes are attributable to family formation, education and occupation, we use the decomposition introduced by Gelbach (2016). In our setting, this involves comparing the coefficient on the feminine conformity index in our saturated regressions (equation (1) in the main text, reproduced in equation (2) below) to the coefficient on the feminine conformity index in a regression which additionally includes our mediators (equation (3)):

$$y_i = \beta_0 + \beta_1 \text{FCI}_i + \beta_2 \text{Cog11}_i + \beta_3 \text{Intern11}_i + \beta_4 \text{Extern11}_i + \gamma' \mathbf{X_i} + u_i \tag{2}$$

$$y_i = \tilde{\beta}_0 + \tilde{\beta}_1 \text{FCI}_i + \tilde{\beta}_2 \text{Cog11}_i + \beta_3 \text{Intern11}_i + \beta_4 \text{Extern11}_i + \tilde{\gamma}' \mathbf{X_i} + \tilde{\delta}' \mathbf{M}_i + e_i \tag{3}$$

where y_i is the outcome of interest, FCI is our feminine 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 selected mediating variables (years of schooling, marital status and number of children, and occupation).

The difference between the two coefficients on the feminine conformity index, $\beta_1 - \tilde{\beta}_1$, is the portion of association between feminine 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:

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

where $\kappa_{M_j}^{FCI}$ is the partial coefficient from a regression of the jth element of $\mathbf{M}_i$ on the feminine conformity index (as well as our set of age-11 covariates), and $\tilde{\delta}_{M_j}$ is the jth element of the coefficient vector $\tilde{\delta}$. The share of the association between feminine conformity index and the outcome explained by the jth mediator is thus given by $\kappa_{M_j}^{FCI}\tilde{\delta}_{M_j}/(\beta_1 - \tilde{\beta}_1)$. 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 (2), equation (3) additionally contains missing dummies for the mediators. Because these missing dummies explain a small fraction of the association between the feminine conformity index and outcomes, the total share explained by education, family formation and occupation in the last column of Table 8 does not exactly equal the difference $\beta_1 - \tilde{\beta}_1$.

E A Model of Feminine Conformity

Our empirical results show that feminine conformity predicts women's labor market outcomes. However, we have so far been agnostic about *why* feminine conformity influences these outcomes. This appendix presents a simple model to show two potential interpretations of why feminine conformity may affect economic outcomes, which are consistent with our empirical findings.

A first interpretation of our feminine conformity index is that it measures a girl's preference to engage in activities and behaviors which are typically associated with women in a society. For example, the index might capture heterogeneity in utility derived from home produced goods. A second interpretation of our index is that it measures real or perceived constraints faced by young girls. For instance, some girls may believe that their labor market return to education are low. Such perceived constraints could prompt these girls to engage in more traditionally female activities, despite not preferring to do so. We present a simple model demonstrating that both interpretations are consistent with the empirical patterns we observe.

E.1 Model Set-up

In our model, women live for two periods - youth and adulthood. In the first period (youth), they choose how much of their time to devote to leisure, education, and work, as well as how much to consume. In the second period (adulthood), the woman chooses how much time to devote to home-production and how much to work (for a wage that is increasing in her education), and how much to consume. In both periods, choices are subject to time and budget constraints. For tractability, we abstract away from occupational and family choices.

Our model has two key parameters which allow for the possibility of heterogeneous preferences and constraints among women. The first (denoted by γ), captures a utility value for home produced goods. 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 feminine conformity could proxy for preferences for home produced goods, especially since much of the index's variation comes from words relating to domestic chores and children (see Section 3.3). To capture heterogeneity in perceived returns to education, we introduce a second parameter ξ which affects the wage that girls expect to face, for a given level of education, such that the returns to education rise in ξ. This allows for the fact that in our model girls may conform with prevailing gender norms because they perceive their returns to education to be low.[33]

We show that a stronger preference for home-produced goods or lower returns to education both imply less education attained during youth. A stronger preference for home-produced goods will additionally mean less time spent in work relative to home-production in adulthood. Both of these implications are consistent with what we find in the data (see Section 4).

[33] An alternative interpretation of perceived constraints is that attaining education has a higher time cost in period 1. We can model ξ as the (heterogeneous) marginal cost of education in the period 1 time constraint - in our model, this would have isomorphic implications to the ones we find with ξ affecting the returns to education.

E.1.1 Period 2: Adulthood

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

$$V_2(ed;\gamma,\xi) \quad = \quad \max_{c_2,hp_2} (1-\gamma)\ln(c_2) + \gamma\ln(hp_2) \tag{4}$$

subject to:

$$\begin{aligned} T &= hp_2 + h_2 \\ c_2 &= h_2 \cdot w_2(ed,\xi) \end{aligned} \tag{5}$$

where ed is the level of education attained in youth. We assume that the utility function is Cobb-Douglas. In adulthood, the wage function is increasing and convex in educational attainment, $w_2(ed,\xi) = w\exp(\xi\cdot ed)$, where w is a constant.

Preferences for home production, proxied by our feminine conformity index, enter through $\gamma \in (0,1)$. This preference parameter drives tastes for home-produced goods relative to market consumption. Perceived returns to educational attainment enter the period 2 maximization problem both directly through period 2 wages, $w_2(ed,\xi)$, as well as indirectly since ξ affects the education decision in period 1.

E.1.2 Period 1: Youth

In period 1, girls choose how much education to attain (ed), how much to work (h_1), and how much time to spend in leisure (l_1). They maximize

$$V_1(\gamma,\xi) \quad = \quad \max_{ed,c_1,l_1} (1-\delta)\ln(c_1) + \delta\ln(l_1) + \beta V_2(ed;\gamma,\xi) \tag{6}$$

subject to

$$
\begin{aligned}
T &= l_1 + ed + h_1 \qquad (7) \\
c_1 &= h_1 \cdot w_1
\end{aligned}
$$

where β is a discount factor. Wages are fixed at w_1. For simplicity, we assume education has no monetary cost. We distinguish between leisure in period 1 and home production in period 2 to highlight that our feminine conformity index captures heterogeneity in home production preferences in adulthood, rather than preferences for leisure enjoyed earlier in life.

E.2 Model Implications

We first study how preferences and constraints (parameterized by γ and ξ, respectively) provide alternative interpretations of the observed relationship between feminine conformity and educational attainment. Solving the period 1 problem in (6) yields the following optimality condition:

$$
\beta \frac{\partial V_2(ed; \gamma, \xi)}{\partial ed} = \frac{\partial u(c_1, l_1)}{\partial c_1} w_1 \qquad (8)
$$

which shows 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, Appendix E.3 shows that we can write equation (8) as:

$$
\beta(1-\gamma)\xi = \frac{(1-\delta)w_1}{c_1} \qquad (9)
$$

Using the time budget constraint to solve for period 1 consumption in the above equation and rearranging yields:

$$
ed = T - \frac{1}{\beta(1-\gamma)\xi} \qquad (10)
$$

The term $\frac{1}{\beta(1-\gamma)\xi}$ depends on the gains from spending time in leisure in period 1 relative to the gains from education enjoyed through higher consumption in period 2 (captured by $\beta(1-\gamma)\xi$). The chosen educational attainment is falling in the preference for home production (γ), and is increasing in the perceived returns to education (ξ).

Appendix E.3 shows that the first-order conditions from the optimization problem in adulthood yields:

$$hp_2 = \gamma T, \tag{11}$$

and thus home production is proportional to γ, which is as expected based on the Cobb Douglas utility formulation. Women with stronger preferences for home production (higher γ) spend a higher proportion of their time in home-production, and as a consequence, consume less. Note that the return to education parameter ξ does not affect home production in period 2 since $w_2(ed, \xi)$ only affects consumption. This is a straightforward implication of Cobb Douglas preferences: income and substitution effects of a change in wages perfectly offset. If substitution effects dominate (and using CES preferences), it is possible to show that home production time is falling in the returns to education.

Our simple model matches several key patterns in the data. The first is that girls with more feminine conformity – that is, girls with a stronger preference for home-produced goods or a lower perceived return to education – attain less education. This is what we show in Table 7. When we model higher feminine conformity as stronger preferences for home production, we gain two additional insights which match our data. The first is that women with higher feminine conformity tend to work fewer hours over their life cycle (a result we show in Table 4). Furthermore, they are found to spend more time in home production activities, which is consistent with our results in Table 7 showing that our index predicts marriage and childbearing at a younger age (which arguably represents spending a greater share of adult life in home production).

Our model structure implies that some, although not all, of our insights are isomorphic to whether we interpret our index of feminine conformity to be a measure of preferences or (perceived) return to education.

E.3 Model derivations

The constrained maximization problem in equations (4) and (5) can be rewritten in Lagrangean form as:

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

The first-order conditions are:

$$\frac{(1-\gamma)}{c_2} = \mu$$

$$\frac{\gamma}{hp_2} = \mu w_2(ed, \xi)$$

Combining these conditions with the budget constraint gives and solving yields:

$$
\begin{aligned}
c_2 &= (1-\gamma) w_2(ed, \xi) T \\
\Rightarrow hp_2 &= \gamma T.
\end{aligned}
$$

Turning to period 1 (youth), after solving out for consumption in equations (6) and (7) , the maximization problem can be rewritten as

$$V_1(\gamma, \xi) = \max_{ed, l_1} (1-\delta)\ln((T - l_1 - ed) w_1) + \delta \ln(l_1) + \beta V_2(ed; \gamma, \xi)$$

The first-order conditions are:

$$FOC_{l_1} : -\frac{(1-\delta) w_1}{c_1} + \frac{\delta}{l_1} = 0$$

$$FOC_{ed} : \frac{-(1-\delta) w_1}{c_1} + \beta \frac{\partial V_2(ed; \gamma, \xi)}{\partial ed} = 0$$

The second FOC can be rewritten as:

$$\beta\frac{\partial V_2(ed;\gamma,\xi)}{\partial ed} = \frac{\partial u(c_1,l_1)}{\partial c_1}w_1 \tag{12}$$

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. This is equation (8) of Appendix E.2.

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

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

Using this equation, we rewrite Equation (12) as

$$\begin{aligned}
\frac{(1-\delta)w_1}{c_1} &= \beta(1-\gamma)\xi \\
\Rightarrow \frac{(1-\delta)}{T-l_1-ed} &= \beta(1-\gamma)\xi.
\end{aligned}$$

Noting that $l_1^* = \delta(T-ed)$ in period 1, we can write the final expression as:

$$\begin{aligned}
\frac{(1-\delta)}{T-\delta(T-ed)-ed} &= \beta(1-\gamma)\xi \\
\frac{1}{T-ed} &= \beta(1-\gamma)\xi \\
\Rightarrow ed &= T - \frac{1}{\beta(1-\gamma)\xi}
\end{aligned}$$

This is Equation (10) of Appendix E.2.