University of St Andrews | Business School

Economics Discussion Papers

Childhood Aspirations and Adult Outcomes

Margaret Leighton and Irina Merkurieva

Economics Discussion Paper No. 2505
Published: 17 December 2025
JEL classification: J24; J26; Z13
Keywords: Aspirations; Education; Natural Language Processing; NCDS

Department of Economics, Business School, University of St Andrews
Discussion Papers Series
ISSN 2978-4026
https://ideas.repec.org/s/san/econdp.html
bschoolecondor@st-andrews.ac.uk

Childhood Aspirations and Adult Outcomes*

Margaret Leighton[†]

Irina Merkurieva[‡]

This version: December 17, 2025

This paper extracts aspirations from texts written in childhood by members of a British longitudinal cohort, and explores how these relate to later life outcomes. Applying Natural Language Processing (NLP) tools to short essays collected at age 11, we identify four aspiration themes: family, hobbies, financial success, and career. The weight of these four themes varies substantially across respondents, with girls on average placing more weight on family, and boys on financial success. Aspirations extracted using our method are strongly predictive of later life outcomes, even when controlling for detailed measures of early life environment, ability, and family background. These associations are often highly heterogenous by gender; for example, family-related aspirations are associated with higher educational attainment for men, but lower educational attainment for women.

Keywords: aspirations, education, natural language processing, NCDS
JEL classification: J24, J26, Z13

*We are grateful to Emma Tominey, seminar participants at the University of St Andrews and the FourSaNE Applied Microeconomics conference for helpful comments and suggestions. We thank Georgi Butch, Zachary Halna du Fretay, Matilda Giaever-Enger, Leonardo Dhinsa, Nina Bucekova, Andrea Moutzouri and Lariena Yang for research assistance. The analyses in this work are based wholly or in part on analysis of data from the 1958 National Child Development Study (NCDS). The data was deposited at the UK Data Archive by the Centre for Longitudinal Studies at the Institute of Education, University of London. NCDS is funded by the Economic and Social Research Council (ESRC).

[†]Department of Economics, University of St Andrews Business School, Castlecliffe, The Scores, St Andrews, UK, KY169AR. E-mail: mal22@st-andrews.ac.uk.

[‡]Department of Economics, University of St Andrews Business School, Castlecliffe, The Scores, St Andrews, UK, KY169AR. E-mail: i.merkurieva@st-andrews.ac.uk.

1 Introduction

Children with similar early achievement scores often go on to have very different careers. But what sets high achievers apart from the rest? A growing body of research identifies aspirations formed in childhood as an important determinant of later-life outcomes (Fruttero et al., 2024, Lybbert and Wydick, 2018, Favara, 2017). While there are compelling theoretical arguments to support this, and a nascent body of empirical work building on the theory, two empirical challenges must be reckoned with: measurement and follow-up.

The challenge of measuring aspirations is faced acutely by researchers collecting primary data. In a recent review of the empirical literature on aspirations, Fruttero et al. (2024) identify 28 different measures used across 30 papers. Yet even when reliable measures are at hand, investigation of the longer-term implications of aspirations is severely limited by the linearity of time: children surveyed today will not complete their careers for many decades.

In this paper, we develop a methodology for extracting information about aspirations from free-text statements. Our approach employs Latent Dirichlet Allocation (LDA), an unsupervised topic modelling technique in Natural Language Processing (NLP) used to discover underlying topics in large collections of natural language text (Blei et al., 2003). Given an appropriate set of inputs, our method generates low-dimensional measures of aspirations which are suitable for use in econometric analysis.

We apply our method to the free-text essays collected as part of the National Child Development Study (NCDS), an ongoing longitudinal study which tracks UK children born in a single week of 1958. The NCDS is exceptional for its unparalleled longitudinal coverage, following individuals from birth into their sixties, and for the extraordinary richness of its individual-level information collected throughout this period. Our focal dataset are 10,506 short essays written by NCDS respondents at the age of 11, in response to the prompt to 'imagine your life at age 25.' These essays do not directly report aspirations, but contain a wealth of information about what the responding child imagines for themselves in their adult life.

Our method identifies four key aspiration themes in the body of essays: family, hobbies, career, and financial success. Each theme's relative importance for the writer is described by a continuous weight, allowing a nuanced measure aspirations across a range of life priorities. We evaluate the relevance of these measures by estimating their power to predict career choices over the lifetime.

Conditional on child's background at age 11, we find that our four aspiration measures are strongly associated with later educational and family outcomes, with clear gender differences. For girls, stronger career aspirations predict longer participation in education and a higher likelihood of university attainment, while hobbies and leisure-oriented aspirations are linked to delayed family formation and smaller family size. Among boys, the patterns are weaker and often reversed, with family-oriented aspirations associated with longer schooling and higher chances of obtaining a degree. Overall, the results suggest that early-life aspirations shape both educational trajectories and key family decisions well into adulthood.

This paper makes three contributions to the literature. First, by investigating the association between aspirations and economic outcomes, we provide new evidence on the long-term influence of aspirations formed in childhood. In their recent survey paper, Fruttero et al. (2024) find strong evidence for an association between aspirations and a range of outcomes. Our work contributes to this by corroborating these trends in exceptionally rich data from a representative, longitudinal cohort in the UK. In contrast to studies that focus on the formation of educational aspirations during adolescence and the socio-economic gradient therein (e.g., Agasisti and Maragkou, 2023), we are able to trace aspirations measured earlier in life and link them to realised outcomes over the life course. Our empirical analysis complements a theoretical literature that models aspirations as reference points shaping behaviour and investment decisions (Besley, 2016, Dalton et al., 2016, Genicot and Ray, 2017), as well as experimental and quasi-experimental evidence highlighting the role of aspirations in educational choice (Page et al., 2007). By providing longitudinal evidence on the persistence and predictive power of early-life aspirations, our paper helps bridge the gap between these theoretical and experimental insights and long-run observed economic outcomes.

Second, we provide an accessible method for extracting themes from free text. While none of the techniques we use here are new in themselves, this paper provides step-by-step guidance for applying them. This methodological contribution can make the application of these methods more accessible to economists or other social scientist with minimal experience using NLP tools.

Finally, our specific application and the data it generates extend the scope of the NCDS, one of the 'work-horse' datasets for understanding long-term individual outcomes in the UK. Prior to their recent digitisation (Goodman et al., 2017), analysis of NCDS childhood essays was limited to a subsample of 495 which had been manually recorded for topics and themes (Elliott et al., 2007, Morrow and Elliott, 2021). Newly-accessible artificial intelligence tools such as NLP allow us to draw upon an objective in-depth examination of the entire collection of NCDS essays. Our work joins other recent and ongoing studies applying such techniques, such as Ayyar et al. (2024)'s study of gender conformity and later-life outcomes and Laurin et al. (2024)'s work on social class.

The remainder of the paper proceeds as follows. Section 2 describes the NCDS dataset and its key features exploited in this paper. Section 3 provides a step-by-step explanation of the LDA methodology, outlines its application to the NCDS essays, and presents the main results of the topic modelling. In Section 4, we document the novelty of our measures of aspirations by correlating them with respondents' own and family characteristics measured in childhood. In Section 5 we demonstrate the predictive power of aspirations by exploring their associations with later life outcomes, conditional on other characteristics. Section 6 concludes.

2 Data

We use data from the National Child Development Study (NCDS), 1958-2025[1]. The NCDS is a longitudinal birth cohort study that follows individuals born across England, Scotland and Wales in a single week between 3 and 9 March 1958. The initial sample consisted of 17,415 newborns, representing 98% of all babies born in the country during these days. Over time, it was expanded to include individuals born abroad during the reference week who later moved to live in the UK. The sample was last updated in 1974, when respondents turned 16, bringing the total cohort size to 18,558.

The survey tracks life histories of the surviving members of this sample who continue to reside in Great Britain. The main body of currently available data were collected through eleven major follow-up surveys (sweeps) when the cohort members were 7, 11, 16, 23, 33, 42, 46, 50, 55, and 62 years old. The NCDS is remarkable for the wide range of information it collects on social, economic, educational and health outcomes. In particular, it offers an extensive set of childhood background and development measures gathered from children as well as their parents and teachers, including cognitive and non-cognitive skills, test scores, family background and circumstances, school achievement, and life orientation. Beginning at age 23, respondents were interviewed directly, providing detailed information on their educational attainment (including both level and field of study), health, labor market and family histories.

At the age of 11, NCDS cohort members were asked to write short essays providing open-ended reflections on their lives, future plans, and expectations looking fourteen years ahead. The essays were written in 1969, when cohort members were in their final year of primary school. Participants had 30 minutes during school hours to complete their responses. The answers, originally handwritten, were subsequently transcribed and anonymised.

The essay prompt was formulated as follows:

[1]University College London, UCL Social Research Institute, Centre for Longitudinal Studies (2024).

"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 overall response rates for the age 11 data collection sweep was 92% of all cohort members known to be alive and residing in the UK at the time of data collection (the target sample). A total of 82% of this target sample participated in the essay-writing exercise. After excluding blank and non-informative entries, as well as 2,990 missing or lost essays, usable texts are available for 77% of the essay writers. Relative to the target sample, this is a response rate of 63%. The final essay sample consists of 10,506 texts, with an average length of 197 words each.

Table 1 presents response rates and selected descriptive statistics for three samples: the set of respondents with available essays; all NCDS participants interviewed during the essay writing sweep; and the target sample. The characteristics compared in Table 1 are drawn from the data recorded at birth, which is available for the full target sample. At age 11, the sample of essay writers appears quite similar to both the interviewed and target samples. In our data analysis, we follow Silverwood et al. (2024) to address the issue of data attrition in the later NCDS sweeps.

We use Age 11 essays to extract childhood aspirations and gain insights into children's visions of the future. Given the open-ended nature of the question prompt, we expect responses to reflect genuine life goals and aspirations. Since the essays were written before major educational or career decisions were made, they are less likely to be influenced by self-assessment biases or retrospective rationalisation. While children's responses would naturally be shaped by their environments and prevailing social norms, they were not yet subject to the adjustment processes that come with realised life events.

Figure 1 provides and example essay to illustrate the range of aspirations a child might express. This is a generated example, constructed to reflect the style, content, and writing ability encountered in the NCDS essay collection. In this case, the writer emphasises entering paid work immediately in order to achieve financial security, with ambitions centred on stable manual employment. Alongside these economic goals, the essay highlights strong family-oriented aspirations, including marriage, raising children, and supporting parents later in life. The child also imagines a domestic life, with home ownership, keeping animals, and leisure activities such as football and fishing.

3 Latent Dirichlet Allocation (LDA)

To extract and summarise relevant information about aspirations from the essay texts, we employ the Latent Dirichlet Allocation (LDA), a widely used Natural Language Processing (NLP) topic modelling technique. LDA is designed to uncover underlying topics in large collections of natural language text (Blei et al., 2003). It is a generative probabilistic model that applies an unsupervised learning approach, identifying topics within text collections without requiring prior data labelling. This makes LDA particularly effective for large-scale analysis of unstructured texts without imposing researcher bias.

The core assumption of LDA is that each document in a collection consists of a mixture of different topics. Each topic is characterised by a distinct distribution of words, and each word in the document collection is associated with at least one topic. Based on this assumption, the algorithm uses statistical inference to determine the most likely distributions of words across topics and topics across documents. The output consists of a predefined number of topics, each represented by a list of words with corresponding weights indicating their importance to the topic. These results can then be used to infer the distribution of topics within documents and assess the relative importance of each topic in a given text.

Applying LDA to the essays in the NCDS collection involves several preprocessing and modelling steps. We outline these steps below and explain how this approach was used to extract information about aspirations. The analysis was conducted in Python using several open-source libraries: Gensim for topic modelling (Řehůřek and Sojka, 2010), SpaCy for tokenisation and lemmatisation (Honnibal et al., 2020), and the Natural Language Toolkit (NLTK) for removal of stop words (Bird et al., 2009).

3.1 Text cleaning

The essay texts have been transcribed as accurately as possible to match the originals, preserving any mistakes and spelling errors. These errors are highly prevalent in the Age 11 essays, which were written by the children themselves. Transcribers used asterisks to replace illegible letters and mark partially unclear words that were substituted with their best guesses, as seen in the sample essay reproduced in Figure 1.

To maximise the amount of information that can be effectively used in the analysis of essays, we first correct spelling mistakes and reconstruct unreadable passages. This step is crucial because misspelled words may either be unique and thus uninformative for identification of thematic patterns, or contain contextual errors that distort the intended meaning of the texts. For example, the description of jobs in the sample essay includes non-existent word 'fctory' instead of 'factory' and contextually inappropriate 'cars' instead of 'cats'.

While some spelling errors are straightforward and fairly easy to fix, others—such as illegible and out-of-context words—can present substantial difficulties. To automate this process, we use context-aware spell-checking based on the GPT4All open-source platform and Nous-Hermes 2 model running on Mistral 7B DPO (Anand et al., 2023, Teknium et al., 2025). The first column in Table 2 shows how this procedure generates a clean semantically coherent text from the original corrupted sample essay. To ensure the quality of the automated corrections, research assistants manually checked the spelling and coherence of a sample of essays, confirming that the procedure produced reliable results.

Spelling correction significantly enhances the quality of essay texts and improves the results of data analysis. *Hapax legomena*—words appearing only one time in the collection of texts—are not helpful in topic modelling: LDA relies on both word context and the likelihood of joint occurrence of patterns in the text. The Age 11 essays initially have 34,165 hapax legomena (over 60% of the words that appear in the texts occur only one time): after spellchecking, this drops to 7,548.

3.2 Preprocessing & dictionary formation

The Age 11 essay collection contains a total of 2,041,020 words, featuring 50,055 distinct entries. For use in the LDA model, the texts of analysed essays must be transformed into bags of words (BoW). This involves several preprocessing steps. First, each document is converted into a list of tokens—meaningful text units that in our application are defined as individual words. Tokenisation transforms texts into lower case and removes all punctuation, symbols and numbers.

The second step is lemmatisation, a process of reducing words to their dictionary forms (lemma). Lemmatisation facilitates recognition of meaning in different word forms based on their part of speech. For example, 'go', 'went' and 'going' are all lemmatised as 'go' and subsequently recognised as the same word by the model. We restrict lemmatised texts to contain only nouns, verbs and adjectives. Column 2 in Table 2 shows the results of tokenisation and lemmatisation steps for the sample essay.

Next, we remove *stop words*—frequently used tokens that have limited semantic value on their own. In a small-scale model like ours, stop words introduce noise and make it harder to define meaningful topics. We use a full list of 198 standard English stop words from the NLTK library. In addition, we define 61 custom stop words that seemed to contribute to noisy outcomes during initial modelling steps. For instance, we removed the word 'life' which featured in the prompt and the phrase 'illegible fiche' which appeared in many essays as a result of procedures used in the early stage of text digitisation.

Finally, we construct and apply bigram models to capture the joint meaning of words that frequently occur in pairs. We identify bigrams using Gensim's Phrases model, which detects statistically significant multi-word expressions in lemmatised text; we require a minimum co-occurrence frequency of five and set a conservative scoring threshold of 100 to retain only strongly associated word pairs. The learned phrase models are then applied using Phraser transformations. Examples of bigrams in NCDS essays include 'air hostess' and 'swimming pool'. Incorporating bigrams improves LDA performance by reducing lexical ambiguity and producing more semantically coherent and interpretable topics. The final result of the pre-processing steps is shown in Column 3 of Table 2.

With preprocessing now complete, further analysis of text data requires a way to identify and distinguish individual tokens. This is done by forming a dictionary, which links each individual token from preprocessed documents to a unique integer identifier. The dictionary does not include words that are unlikely to be informative for identification of topics, either because they are unique or because they are very common. Our dictionary excludes the 6,387 words that only appear in a single document (e.g., ‘juniper’ and ‘wholegrainer’) as well as common words that appear in more than 50% of the documents (e.g. the word ‘work’).

The resulting dictionary contains 8,319 words. It provides the foundation for the *corpus*, which in the context of an LDA model is a collection of vectorised document representations. Each word in the document is replaced by its dictionary identifier along with its number of occurrences in the document. Column 4 in Table 2 shows a corpus entry obtained from the sample essay. Together, the dictionary enables the model to interpret and map topics back to actual words, while the corpus provides the numerical input the model uses to infer topic distributions.

3.3 Topic modelling

The Gensim implementation of the LDA algorithm requires the user to specify several computational parameters, including the number of topics to extract. We perform a grid search over these parameters to maximise the model's topic coherence score. Coherence measures the semantic consistency of the top words within each topic and is computed using cosine similarity between word vectors, yielding values in the range from 0 to 1. Higher coherence scores suggest better semantic consistency of topics, with conceptually related top words more likely to appear together in the corpus documents. As a general guideline, a coherence score value above 0.6 implies highly coherent topics that are interpretable for humans (see the review in Röder et al., 2015). We measure coherence using Gensim's *c_v* coherence metric.

We find the number of topics to be the most impactful parameter for the model's coherence score. Figure 2 shows how, after initial rapid improvement in coherence with increased number of topics, coherence stabilises and peaks between 10 and 20 topics. Our final model has 13 topics and a coherence score of 0.62. For other model parameters, we set the number of passes through the corpus during training at 500, the maximum number of iterations through the corpus for each pass at 10, the chunk size for the number of simultaneously processed documents at 2000; the Dirichlet asymmetry and topic-word priors, α and η, are learnt from the data during training.

The primary output of an LDA topic model is the probability distribution of dictionary words within topics. Words that are assigned the highest probability are the most informative of the topic contents, and each topic can be seen as a collection of dominant keywords. To demonstrate these results in our model, Table 3 lists the ten most important words and their weights for each of the 13 extracted topics. In a good model, it should be easy to infer the topic contents from its keywords, which appears true for our topics. For example, the dominant theme in Topic 2 is football and sports: the first five words all refer to the game and collectively account for a third of the topic contents. Topic 11 is about family: the top words 'child', 'husband', 'boy' and 'girl' represent over 25% of its contents.

Another way to assess the quality of the topic model is by visual analysis of the topic map (Figure 3). Each bubble on the map represents a topic identified by the model, with the bubble size corresponding to the topic prevalence in the corpus. The position of bubbles on the plane is determined by principal component analysis, and the distance between bubbles reflects the similarity between topics. Closer bubbles indicate more similar topics. In our model, the most important topics are related to work and money (Topic 10) and family (Topic 11). An intuitive example are slightly overlapping bubbles that correspond to Topics 4 and 9, both featuring animals (pets and horses, respectively). Our topic map has fairly large non-overlapping bubbles scattered over the entire plain, suggesting a good model with distinct topics that cover diverse aspects of the corpus.

Overall, our LDA model demonstrates strong performance across multiple evaluation criteria. The post-estimation analysis affirms that we have built a robust model with a high coherence score and interpretable topics that provides a good representation of the underlying thematic structures in the essays.

3.4 Topic analysis

Once the model is trained, we can use it to analyse individual essays and compute the percentage of each document's text associated with particular topics. In most cases, multiple topics will be present in a single document. For example, the last column of Table 2 shows that the sample essay is dominated by five topics jointly capturing 80% of its contents (topics 10, 4, 11, 2 and 12).

Comparing a topic's keywords and the essay contents, it is straightforward to understand the intuition behind weight assignment. For example, Topic 10 (54%) captures the respondent's intention to find a good job and make a lot of money, Topics 2 and 4 (15%) relate to the hobbies (pets and football), while Topic 11 (9%) reflects the wish to have a family. The percentages indicate the proportion of words in the essay that are most strongly associated with each topic according to the LDA model, reflecting the relative emphasis of each theme within the essay text.

Overall, the topics identified by the algorithm and their assigned weights provide a compelling summary of the essay's contents. However, the number of topics is quite large for analysis. Furthermore, several topics detected by LDA share overall subject themes which, although easily interpretable by a researcher, may not be possible for an algorithm to pick up without preliminary training. In the sample essay, this is seen in Topics 2 and 4 that jointly capture hobbies and leisure activities. To make further analysis more tractable, we manually consolidate detected topics into four overarching themes: family, hobbies, career, and financial wellbeing.

The last column of Table 3 summarises the assignment of topics to themes. The family theme corresponds to Topics 3 and 11, which are focussed on children, marriage and home life. The theme of hobbies and leisure includes Topics 2, 4, 9 and 13 which all talk about various interests beyond raising a family and having a job (among others, football, riding and reading). The career theme incorporates Topics 1, 5-7 and 12: these all include specific jobs and occupations (including, for example, nursing, teaching, and bus driving). Finally, the theme of financial wellbeing combines Topics 8 and 10, where the main emphasis is on making money and owning material items, such as a house or a car.

We define a theme weight as the summary weight of all topics that make up that theme. When viewed from this prospective, the sample essay reflects a life vision with financial wellbeing taking the top place (58%), followed by hobbies (20%), family (12%) and career (10%). The distribution of themes across essays reveals substantial differences in respondents' priorities. Financial wellbeing emerges as the most important theme with 30.5% average weight. Other themes are also well-represented, with the lowest weight given to family and home life (15.6%). Theme weights have a wide spread over the unit interval, suggesting that while most respondents discussed a variety of interests in their essays, some focussed instead on a single thematic area. The most common dominant themes in such essays are career and financial wellbeing: each of these themes contributes over 50% in at least 10% of essays in the sample. Table 4 summarises descriptive statistics for theme weights.

3.5 Theme analysis

We interpret these four themes as capturing the aspirations of the respondents: tasked with describing their future lives, these are the topics that came to mind. These themes differ systematically across two more fundamental dimensions: work-life balance and conformity to social norms. In the first dimension, career and money concerns represent a higher weight on work versus leisure time, while a greater interest in hobbies and family is associated with a stronger preference for non-work time. On the other hand, family and career aspirations exhibit strong alignment with socially-approved ways of life; a dominant focus on earning money or hobbies conforms less with social norms.

To explore the variation in respondents aspirations across these two dimensions, we convert aspiration theme weights W and plot in in two dimensions, defining the horizontal coordinate along the work-life balance axis as

$$x = [W(family) + W(hobbies)] - [W(money) + W(career)]$$

and the vertical coordinate along the conformity axis as

$$y = [W(family) + W(career)] - [W(hobbies) + W(money)].$$

Figure 4 shows the distribution of individual respondents in the plane defined by these two dimensions. While the cloud of responses is skewed towards work on the work-life axis, there is a wide spread of weights across the two dimensions. The sample essay, marked by point O at coordinates $(-0.36, -0.56)$, falls in the third quadrant of the plane. This position indicates that the writer places greater emphasis on work than leisure on the work-life balance axis (negative horizontal value) and aligns less with socially approved norms on the conformity axis (negative vertical value).

We hypothesise that these two dimensions capture fundamental aspects of individual preferences. Specifically, we expect the work-life balance to map into consumption-leisure weights and the degree of conformity into the risk aversion. These two parameters driving individual decision-making in a life cycle model of labour supply and consumption decisions. If this is true, we would expect aspirations expressed in the Age 11 essays to correlate with key life decision and outcomes, such as education, employment and savings. In the next two sections, we first check the novelty of our aspiration themes with respect to existing data; we then explore the relationship between aspirations and later outcomes.

4 Aspirations and childhood background

Aspirations are shaped by a child's home and social environment. Do the aspiration themes we have identified in the Age 11 essays simply reflect the respondent's personal characteristics and upbringing—characteristics already extensively documented in these data—or do they capture something else? In other words, is there new information in these themes that might be useful for understanding the respondents and modelling their life course?

To investigate this, we explore how our four themes correlate with a broad range of NCDS variables describing children and their early-life backgrounds. If these characteristics are strongly predictive of aspiration themes, it suggests that the LDA procedure has extracted meaningful data from the responses; however, it also suggests that these aspiration themes have little empirical value-added. Limited correlations, on the other hand, suggest that the aspirations are capturing novel characteristics not already in the quantitative data. In such a case, however, further validation is needed to demonstrate that these themes are meaningful.

The NCDS provides a rich set of variables that describe early childhood conditions for cohort members. Our choice of variables is guided both by existing research on predictors of later-life outcomes and by work from the NCDS team, which used machine learning techniques to identify survey variables that best predicted individual-level attrition in later sweeps (Silverwood et al., 2024). We retain as independent variables a range of demographic and social characteristics, including gender, cognitive ability, parental interest in schooling and expectations for the child's future, father's social class and mother's demographic characteristics (age, marital status and education), as well as household crowding (the number of people per room). Measures of ability, parental interest, and expectations are derived from multiple relevant survey items using principal component analysis and are included in standardised form.

Table 5 presents the results of this analysis, with one regression for each aspiration theme (measured as a percentage weight of the respondent's essay). The first take-away from this table is that, with the exception of gender, the aspiration themes are largely uncorrelated with individual characteristics. While there are some statistically significant associations, these are small; for the most part, the associations are both statistically insignificant and trivial in magnitude.

Gender, the exception to this general trend, emerges as the most important predictor across all four themes. Females place greater emphasis on the two themes associated with the high conformity dimension: family (+11.7%, approximately 1 SD) and career (+3.4%, 0.2 SD), while giving less weight to financial success (-11.5%, 0.7 SD) and hobbies (-3.7%, 0.2 SD).

Compared with the other themes, career-related aspirations are associated with a slightly broader set of factors: they are stronger among children whose parents hold higher expectations, those raised in more crowded households, and lower among children with higher verbal ability. Beyond verbal ability (-2.4%, 0.15 SD), however, even these statistically significant associations are small.

Beyond gender, we observe only occasional, weak, and small-magnitude correlations. Among the three ability measures, verbal ability shows the most consistent associations: a 1 SD increase is linked to a 2.4% decrease in career aspirations and a positive association with the other themes. Math ability is positively associated with hobbies and negatively with family-related aspirations. Parental interest in schooling does not show a significant relationship with any aspiration theme, while a 1 SD increase in parental expectations is associated with a 1% increase in career aspirations and reduced emphasis on other themes.

Taken together, these estimates demonstrate that the aspirations themes are capturing characteristics which are distinct from those already described in the quantitive data. They also show that gender is an important predictor of childhood aspirations. But are the aspiration themes we have derived meaningful? Do they have implications for future choices, and later-life outcomes? In the next section we explore how our aspiration measures correlate with later outcomes, controlling for all characteristics which were significantly associated with aspirations in this first stage of validation.

5 Aspirations and later-life outcomes

The NCDS follows cohort members from their birth in 1958 to present. Its rich longitudinal design provides a unique opportunity to examine how childhood aspirations are associated with outcomes throughout their lives. Drawing on data from successive survey sweeps, we investigate how early aspirations are linked to three outcome domains: education, family formation and economic. Each of these outcome areas is addressed in the subsections that follow.

Given the gender differences in aspirations identified in earlier analyses, all models are estimated separately for males and females. We also control for childhood background characteristics which were found to have a statistically significant association with aspirations, so that the estimated relationships between aspirations and later-life outcomes are conditional on early-life achievement and family background. Family aspirations serve as the reference category, and all non-dummy explanatory variables are standardised to allow comparability across determinants of education.

5.1 Education

We consider three educational outcomes: (1) age at which cohort members left full-time education, (2) the likelihood of attaining a university degree, and (3) the expected future earnings of university graduates by subject studied (available only for degree holders). These outcome variables are summarised in Table 6. We measure future earnings using data from the first wave of the Annual Population Survey (APS) conducted in 2004, when members of the NCDS cohort, then aged around 46, were near their peak lifetime earnings. Following the APS classification, degree subjects are grouped into 18 categories, and gross weekly earnings by subject group are used to proxy returns to degree subjects.

We use the log of this value in our regressions.

The results for educational outcomes are summarised in Table 7. Aspirations are meaningful predictors of educational attainment for both genders, although with contrasting patterns of effect: boys with stronger family aspirations stay longer in education compared to boys with other aspirations, while girls display the reverse pattern. This is also true for university degree attainment. For girls, placing even modestly greater emphasis on any type of aspiration other than family substantially increases the likelihood of obtaining a university degree: for example, a 1 SD increase in career aspirations is associated with a 2.7 percentage point increase in the probability of obtaining a degree. For boys the effects are smaller but again reverse those of girls, with e.g. a 1 SD increase in financial aspirations associated with a 1.3 percentage point decrease in the probability of obtaining a degree.

Conditional on attending university, we find no consistent pattern in the choice of subject. However, while family aspirations are negatively associated with the likelihood of university attendance for girls, they have the opposite effect among female graduates: conditional on completing a degree, stronger family aspirations predict choice of degrees with higher expected earnings. For boys the association between degree choice and aspirations are smaller and less consistent.

5.2 Family

We next examine how childhood aspirations relate to family formation and childbearing outcomes. We consider three outcomes related to children: whether the cohort member had any biological children by age 25, the total number of children (including biological, adopted, or fostered) by age 50, and the cohort member's age at the birth of their first biological child. An overview of these outcome variables can be found in Table 6.

The results are summarised in Table 8. For men, we find no meaningful relationship between aspirations and children. For women, however, there are some marked patterns. A 1 SD increase in hobby-related aspirations is associated with a 3 percentage point decrease in the likelihood of having a child by age 25, a 0.4-year increase in the age at first birth, and a reduction of 0.06 in the total number of children by age 50. Financial aspirations display similar, though somewhat weaker, patterns, whereas career aspirations do not differ significantly from family-oriented aspirations in their effects on childbearing.

5.3 Economic outcomes

We consider two sets of economics outcomes: work-related (earnings, hours worked, number of jobs) and asset-related (savings, home ownership, home value). These outcome variables are summarised in Table 6. All outcomes are measured at age 23 (Sweep 4), which maximises sample size by avoiding later-sweep attrition. Although measured at a relatively young age, these indicators provide meaningful insights given the well-documented persistence of income and wealth trajectories over the life course. It should be noted, however, that wealth is measured at the household-level, while all other outcomes we have considered are measured at the individual level - with some respondents already married or partnered by age 23, the values of these variables will be more difficult to interpret.

Table 9 presents the results on employment outcomes: weekly gross earnings in the current job, the number of hours worked per week, and the total number of jobs the cohort member has held up to age 23. While there are no dominant trends apparent in the table, several statistically significant associations are worth mentioning. Weekly earnings are significantly higher for women who expressed financial or career aspirations in childhood, compared with family or hobby aspirations. Point estimates suggest that men with higher financial or family aspirations may have higher weekly earnings than those with hobby or career aspirations, but these differences are not statistically significant. While hours worked and number of jobs are not generally related to our aspirations measures, there is suggestive evidence that women with a strong interest in hobbies work longer hours and change jobs more frequently.

Table 10 presents the results on savings and assets: the total value of savings and investments, a binary indicator of home ownership, and the real value of the primary residence. A strong interest in hobbies is associated with personal wealth and savings behaviour for both genders; however, as with previous outcomes, the effects differ markedly by gender. For women, stronger hobbies-related aspirations are linked to lower wealth accumulation, whereas for men the relationship is positive. Financial aspirations are positively associated with savings and wealth across all three measures, but only among men; while career aspirations have no statistically-significant association with any of these outcomes. While none of the point estimates are statistically significant, it is interesting to note that home value at age 23 is positively associated with family aspirations for women, but negatively associated with this same aspiration among men.

Taken together, we find substantial associations between measured aspirations and later-life outcomes across a range of domains. We draw two conclusions from these results. First, the patterns we observe, and the meaningful variation in these patterns across genders, demonstrate that childhood aspirations are significant predictors of adult socioeconomic circumstances. This is true even after conditioning on early-life background characteristics, which partly determine the formation of aspirations in the first place. Second, these correlations suggest that our method for extracting aspirations is capturing economically meaningful individual characteristics. While machine-learning techniques like LDA are something of a "black box," the aspirations derived through our procedure show meaningful associations with later-life outcomes.

6 Conclusion

Persistent disparities in income and wealth have their roots in the social norms, cultural expectations, and economic structures that shape individuals' opportunities and choices from childhood to retirement. Among these factors, aspirations formed in early childhood may be uniquely malleable. A better understanding of the role that aspirations play in the life course can inform targeted interventions to promote equal opportunities and reduce economic inequality.

We apply a Latent Dirichlet Allocation (LDA) model to the childhood aspiration essays from the NCDS and identify four major aspiration themes: family, hobbies, money, and career. We demonstrate that these themes possess predictive power for later-life outcomes, even after accounting for individual demographics and early-life background characteristics. This novel approach enables us to capture heterogeneity in individual aspirations. Future research can use such data to trace the long-run influence of early aspirations, estimating their importance for educational choices, employment paths, and retirement decisions.

References

Agasisti, T. and K. Maragkou (2023). Socio-economic gaps in educational aspirations: do experiences and attitudes matter? *Education Economics 31*(4), 471–487.

Anand, Y., Z. Nussbaum, B. Duderstadt, B. Schmidt, and A. Mulyar (2023). Gpt4all: Training an assistant-style chatbot with large scale data distillation from gpt-3.5-turbo. `https://github.com/nomic-ai/gpt4all`.

Ayyar, S., U. Bolt, E. French, and C. O'Dea (2024). Imagine your life at 25: Gender conformity and later-life outcomes. NBER Working Paper 32789.

Besley, T. (2016). Aspirations and the political economy of inequality. *Oxford Economic Papers 69*(1), 1–35.

Bird, S., E. Klein, and E. Loper (2009). *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.".

Blei, D. M., A. Y. Ng, and M. I. Jordan (2003). Latent Dirichlet allocation. *Journal of machine learning research 3*, 993–1022.

Dalton, P. S., S. Ghosal, and A. Mani (2016). Poverty and aspirations failure. *The Economic Journal 126*(590), 165–188.

Elliott, J., V. Morrow, et al. (2007). Imagining the future: Preliminary analysis of NCDS essays written by children at age 11. Centre for Longitudinal Studies, Institute of Education, University of London.

Favara, M. (2017). Do Dreams Come True? Aspirations and Educational Attainments of Ethiopian Boys and Girls. *Journal of African Economies 26*(5), 561–583.

Fruttero, A., N. Muller, and O. Calvo-González (2024). The Power and Roots of Aspirations. *The World Bank Research Observer 40*(1), 58–103.

Genicot, G. and D. Ray (2017). Aspirations and inequality. *Econometrica 85*(2), 489–519.

Goodman, A., M. Narayanan, S. Brown, and T. Murphy (2017). Age 11 essays – imagine you are 25, 1969: User guide to the data (first edition). London: UCL Centre for Longitudinal Studies.

Honnibal, M., I. Montani, S. V. Landeghem, and A. Boyd (2020). spaCy: Industrial-strength Natural Language Processing in Python.

Laurin, K., H. R. Engstrom, and M. Huang (2024). What will my life be like when I am 25? How do children's social class contexts predict their imagined and actual futures? *Journal of Social Issues 80*(4), 1433–1459.

Lybbert, T. J. and B. Wydick (2018). Poverty, Aspirations, and the Economics of Hope. *Economic Development and Cultural Change 66*(4), 709–753.

Morrow, V. and J. Elliott (2021). 'I'd keep them tidy': Domesticity, work and nostalgia in girls' imagined futures described in essays written by 11 year olds in 1969. In L. Moran, K. Reilly, and B. Brady (Eds.), *Narrating Childhood with Children and Young People: Diverse Contexts, Methods and Stories of Everyday Life*, pp. 283–306. Springer International Publishing.

Page, L., L. L. Garboua, and C. Montmarquette (2007). Aspiration levels and educational choices: An experimental study. *Economics of Education Review 26*(6), 747–757.

Řehůřek, R. and P. Sojka (2010, May). Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, Valletta, Malta, pp. 45–50. ELRA. `http://is.muni.cz/publication/884893/en`.

Röder, M., A. Both, and A. Hinneburg (2015). Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining*, pp. 399–408.

Silverwood, R., M. Narayanan, B. Dodgeon, M. Katsoulis, and G. Ploubidis (2024). Handling missing data in the CLS cohort studies: User guide. London: UCL Centre for Longitudinal Studies.

Teknium, theemozilla, karan4d, and huemin_art (2025). Nous hermes 2 mistral 7b dpo. `https://huggingface.co/NousResearch/Nous-Hermes-2-Mistral-7B-DPO`.

University College London, UCL Social Research Institute, Centre for Longitudinal Studies (2024). National Child Development Study. [data series]. 14th Release. UK Data Service. SN: 2000032, DOI: http://doi.org/10.5255/UKDA-Series-2000032.

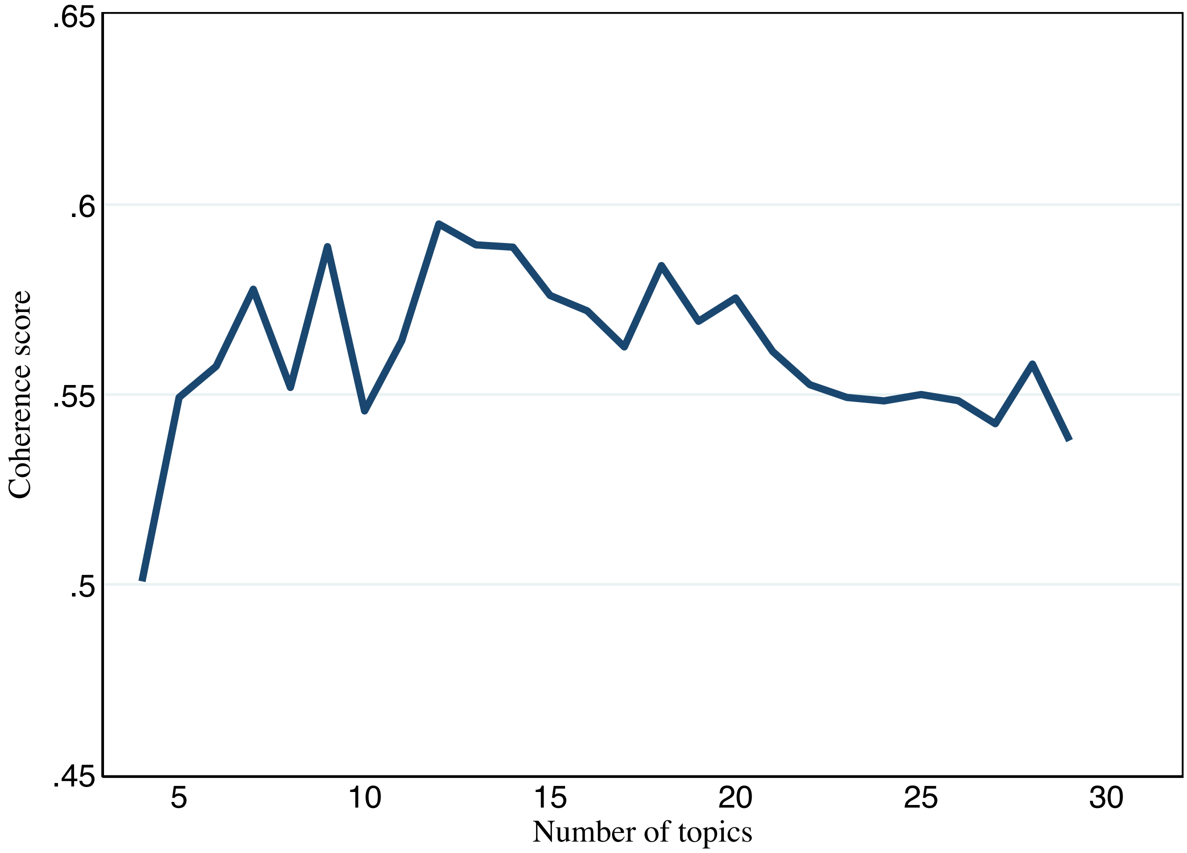
7 Figures and Tables

Figure 1: A sample essay by 11 year old

When I am 25 I think I will go out to work straite away so I can earn my own money. I would like to have a good job, maybe in a fctory or as a mec**nic, so I can be sure I will always have work. When I have saved up enuf money I would like to buy a house with a grden. I would like to keep some animls, like hens cars or a dog, and have a nice place to live. I want to get maried and have chldren of my own and look after them properley. I would also like to help my mum and dad when they get older so they do not have to worry. In my free time I would play football* and go f*shing with my freinds.
Words: 135

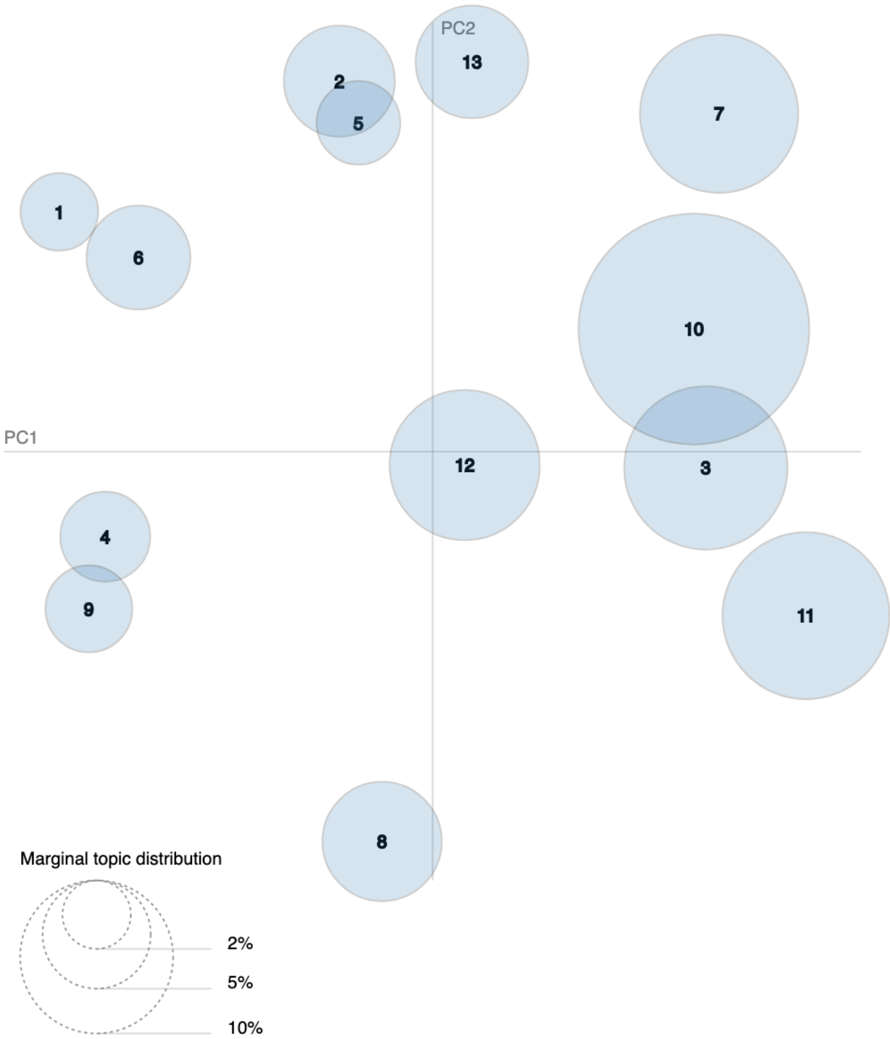
Notes: This essay is a generated example written to reflect the style, themes, and writing abilities represented in the NCDS essays collection. As with the original transcribed essays, it reflects the spelling mistakes present in children's original writing. Asterisks were used by transcribers to replace letters illegible in the original; an asterisk following a word indicates partially legible word replaced with transcriber's best guess.

Figure 2: Coherence score by the number of model topics



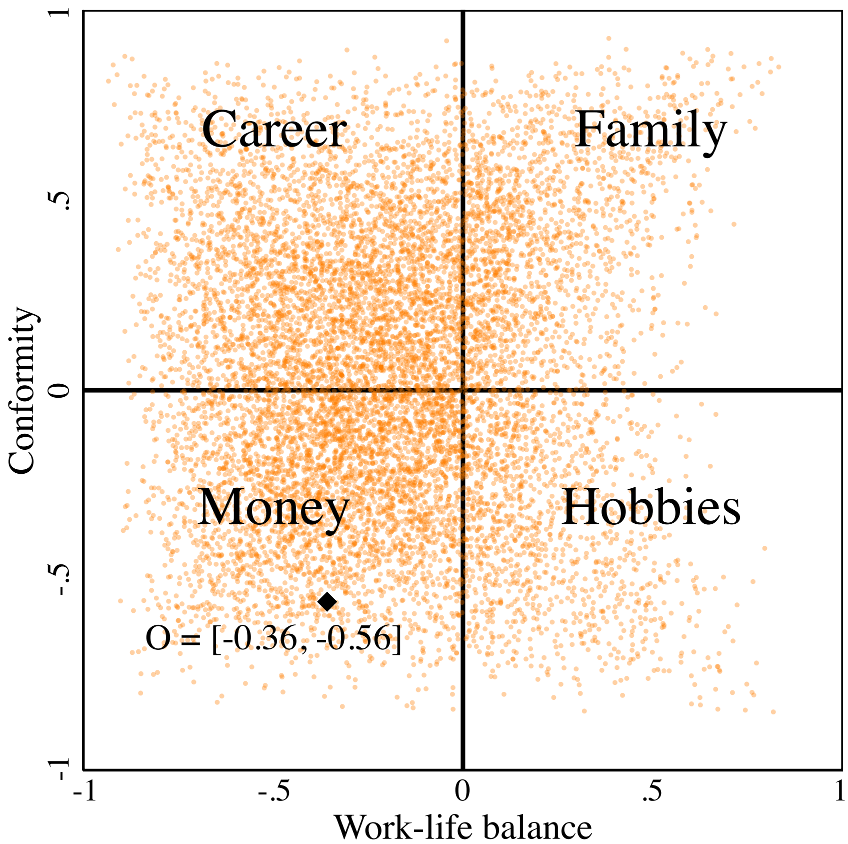
Notes: Coherence computed by Gensim's c_v function. The settings for other parameters of LDA model are 500 corpus passes, 10 iterations, 2000 chunk size; the Dirichlet asymmetry and topic-word priors, α and η, are learnt from the data during training.

Figure 3: Intertopic distance map



Notes: Each circle represents a topic, with size indicating topic prevalence and distance reflecting semantic similarity between topics.

Figure 4: The distribution of aspiration themes



Notes: Point O corresponds to the weights assigned to the sample essay.

Table 1: Essays response rate and descriptive statistics

	Essay writers		Interviewed		Target sample	
Gender, % female	48.4	(50.0)	48.6	(50.0)	48.7	(50.0)
Mother's characteristics:						
Age	27.5	(5.7)	27.5	(5.7)	27.5	(5.7)
Married	0.96	(0.19)	0.96	(0.19)	0.96	(0.19)
Schooling 1 = past minimum)	0.26	(0.44)	0.25	(0.43)	0.25	(0.43)
Father's social class:						
I and II	0.19	(0.39)	0.18	(0.39)	0.19	(0.39)
III and IV manual	0.64	(0.48)	0.64	(0.48)	0.63	(0.48)
III and IV non-manual	0.14	(0.34)	0.13	(0.34)	0.13	(0.34)
V	0.04	(0.20)	0.05	(0.21)	0.05	(0.21)
Living conditions	0.86	(0.61)	0.86	(0.60)	0.86	(0.61)
Sample size (% target sample)	10506	(63)	15337	(92)	16742	(100)

Sample means with standard deviation in parentheses. Characteristics of parents as recorded at birth (Sweep 0).

Table 2: The results of text processing for the sample essay

Corrected essay	Lemmatized text	Non-stop	Corpus entry	Topic weights
When I am 25, I think I will go out to work straight away so I can earn my own money. I would like to have a good job, maybe in a factory or as a mechanic, so I can be sure I will always have work. When I have saved up enough money, I would like to buy a house with a garden. I would like to keep some animals, like hens or a dog, and have a nice place to live. I want to get married and have children of my own and look after them properly. I would also like to help my mum and dad when they get older so they do not have to worry. In my free time, I would play football and go fishing with my friends.	['think', 'go', 'work', 'earn', 'own', 'money', 'like', 'good', 'job', 'factory', 'mechanic', 'sure', 'work', 'save', 'enough', 'money', 'like', 'buy', 'house', 'garden', 'like', 'keep', 'animal', 'hen', 'dog', 'nice', 'place', 'live', 'want', 'get', 'married', 'child', 'own', 'look', 'like', 'help', 'mum', 'dad', 'get', 'old', 'worry', 'free', 'time', 'play', 'football', 'go', 'fish', 'friend']	['work', 'earn', 'money', 'job', 'factory', 'mechanic', 'sure', 'work', 'save', 'enough', 'money', 'buy', 'house', 'garden', 'keep', 'animal', 'hen', 'dog', 'nice', 'married', 'child', 'help', 'mum', 'dad', 'worry', 'free', 'play', 'football', 'fish', 'friend']	[('child', 1), ('garden', 1), ('house', 1), ('married', 1), ('money', 2), ('nice', 1), ('save', 1), ('dog', 1), ('football', 1), ('play', 1), ('factory', 1), ('dad', 1), ('friend', 1), ('job', 1), ('mum', 1), ('buy', 1), ('earn', 1), ('enough', 1), ('help', 1), ('fish', 1), ('mechanic', 1), ('keep', 1), ('animal', 1), ('hen', 1), ('sure', 1), ('free', 1), ('worry', 1)]	(1, 0.0064) (2, 0.0621) (3, 0.0228) (4, 0.0928) (5, 0.0054) (6, 0.0066) (7, 0.0245) (8, 0.0464) (9, 0.0364) (10, 0.5400) (11, 0.0927) (12, 0.0529) (13, 0.0110)
Word count: 135	Token count: 48	Token count: 30	Unique tokens: 27	Composite themes: 12% family; 20% hobbies; 10% career; 58% financial wellbeing.

Table 3: Extracted topics and the distribution of top words

Topic	Top words	Theme
Topic 1	london (0.0807), street (0.0456), number (0.0246), police (0.0218), bank (0.0211), test (0.0206), pass (0.0187), station (0.0187), flat (0.0172), road (0.0103)	3 - Career
Topic 2	play (0.1198), football (0.1131), team (0.0492), club (0.0340), match (0.0262), game (0.0217), footballer (0.0203), cricket (0.0175), sport (0.0127), cup (0.0125)	2 - Hobbies
Topic 3	home (0.0720), bed (0.0450), dinner (0.0384), tea (0.0373), sunday (0.0338), watch (0.0323), saturday (0.0307), shop (0.0250), television (0.0213), afternoon (0.0169)	2 - Hobbies
Topic 4	dog (0.1425), animal (0.0825), pet (0.0475), cat (0.0464), bird (0.0388), vet (0.0203), rabbit (0.0194), fish (0.0139), keep (0.0119), zoo (0.0111)	2 - Hobbies
Topic 5	nurse (0.0875), hospital (0.0714), people (0.0396), doctor (0.0366), help (0.0240), become (0.0197), air_hostess (0.0182), die (0.0177), ward (0.0174), patient (0.0137)	3 - Career
Topic 6	fly (0.0323), army (0.0274), ship (0.0251), plane (0.0250), pilot (0.0178), fire (0.0145), leave (0.0139), boat (0.0133), join (0.0131), navy (0.0126)	3 - Career
Topic 7	school (0.0627), mother (0.0557), friend (0.0381), father (0.0345), enjoy (0.0318), teacher (0.0311), teach (0.0289), sister (0.0252), brother (0.0230), college (0.0184)	3 - Career
Topic 8	garden (0.0699), house (0.0558), room (0.0540), big (0.0421), bedroom (0.0357), kitchen (0.0223), white (0.0215), flower (0.0210), living (0.0161), black (0.0156)	4 - Financial wellbeing
Topic 9	horse (0.0910), ride (0.0787), farm (0.0585), pony (0.0200), stable (0.0184), field (0.0163), race (0.0148), plant (0.0139), cow (0.0134), farmer (0.0118)	2 - Hobbies
Topic 10	job (0.0588), car (0.0462), house (0.0369), money (0.0229), wife (0.0220), buy (0.0218), holiday (0.0168), hope (0.0161), home (0.0156), town (0.0151)	4 - Financial wellbeing
Topic 11	child (0.1771), husband (0.0773), girl (0.0516), boy (0.0454), school (0.0332), marry (0.0298), little (0.0295), baby (0.0263), happy (0.0243), nice (0.0226)	1 - Family
Topic 12	bus (0.0173), man (0.0167), give (0.0137), stop (0.0124), know (0.0122), people (0.0103), way (0.0103), walk (0.0101), break (0.0097), door (0.0097)	3 - Career
Topic 13	book (0.0504), read (0.0428), collect (0.0289), stamp (0.0278), interesting (0.0217), write (0.0173), interest (0.0166), many (0.0157), paint (0.0157), draw (0.0155)	2 - Hobbies

Notes: The first ten words and their weights in the identified topics. The last column shows the assignment of LDA topics to four themes.

Table 4: Theme weights in Age 11 essays, descriptive statistics

Theme	Weight, %						
	Mean	S.D.	Min	10th percentile	Median	90th percentile	Max
Family and home life	15.6	12.0	0.694	2.9	12.4	33.3	73.9
Hobbies and leisure	24.9	15.4	1.442	7.5	21.8	46.6	89.0
Financial wellbeing	30.5	15.5	2.361	11.7	28.5	52.5	87.4
Career and occupations	28.9	16.2	2.629	10.3	26.1	52.4	92.4
Number of essays	10,405						

Table 5: Aspirations by child background

	Aspiration theme weight:			
	Family	Hobbies	Money	Career
Female	11.71***	-3.680***	-11.453***	3.421***
	(0.276)	(0.409)	(0.381)	(0.416)
Math ability	-0.441*	0.771**	-0.049	-0.281
	(0.247)	(0.365)	(0.340)	(0.372)
Verbal ability	1.019***	0.759*	0.641*	-2.419***
	(0.278)	(0.411)	(0.383)	(0.418)
General ability	-0.285	0.016	0.291	-0.023
	(0.262)	(0.387)	(0.361)	(0.394)
Parental interest in school	0.286	0.037	0.075	-0.398
	(0.176)	(0.260)	(0.242)	(0.264)
Parental expectations	-0.340*	-0.087	-0.579**	1.007***
	(0.176)	(0.260)	(0.242)	(0.265)
Father class I and II	-0.527	0.985*	-0.709	0.252
	(0.371)	(0.550)	(0.512)	(0.559)
Mother's age at birth	-0.020	0.014	-0.054	0.059*
	(0.024)	(0.035)	(0.033)	(0.036)
Mother married at birth	-0.884	0.859	-0.315	0.340
	(0.893)	(1.321)	(1.231)	(1.344)
Mother's education, 1 = past min	0.055	0.357	0.322	-0.734
	(0.336)	(0.497)	(0.463)	(0.505)
Number of people per room	-0.022	-0.373	-0.649*	1.044**
	(0.243)	(0.359)	(0.335)	(0.366)
R-squared	0.248	0.018	0.147	0.050
Observations	6345	6345	6345	6345

Outcome variable is the weight (in percentage points) of that aspiration in the respondent's essay: the values of the four aspirations sum to 100. Ability, parental interest and parental expectation variables are standardised to have a mean of zero and a standard deviation of one. OLS estimates with standard errors in parentheses. ***, ** and * indicate statistical significance at 1, 5 and 10% level.

Table 6: Descriptive statistics for later-life outcome variables

	Males				Females			
	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max
Age when left full-time continuous education	17.0	2.0	14	28	16.9	1.8	14	27
University degree holders	0.16	0.36	0	1	0.13	0.34	0	1
Median log-earnings for degree subject	6.22	0.19	5.70	6.54	6.12	0.14	5.70	6.54
Whether had biological children by age 25	0.28	0.45	0	1	0.45	0.50	0	1
Total number of children (latest available)	1.67	1.37	0	9	1.77	1.27	0	16
Age at birth of the first child, years	27.5	5.7	14	50	24.9	5.3	13	44
Gross weekly earnings, Sweep 4	88.9	60.7	0	520	51.4	47.6	0	560
Weekly hours of work at the main job	43.5	10.1	7	96	36.4	8.6	3	96
Number of jobs held by age 23	2.85	1.99	0	9	2.71	1.78	0	9
Total value of savings and investments, age 23	1178.9	4256.6	0	95000	1100.2	4712.9	0	95000
Whether a home owner, age 23	0.22	0.41	0	1	0.36	0.48	0	1
Log house value, age 23	10.25	0.51	6.95	12.21	10.37	0.49	6.73	11.74
Observations	4750				4591			

Table 7: Aspirations and educational outcomes

	I. Age left full-time education		II. University degree		III. Expected degree returns	
	Males	Females	Males	Females	Males	Females
Aspiration theme weights (base category - Family):						
Hobbies	-0.075**	0.026	-0.013*	0.017**	-0.007	-0.022**
	(0.037)	(0.031)	(0.007)	(0.007)	(0.012)	(0.010)
	[0.043]	[0.404]	[0.052]	[0.011]	[0.574]	[0.034]
Money	-0.094**	0.019	-0.013*	0.024***	0.008	-0.026**
	(0.041)	(0.029)	(0.007)	(0.006)	(0.013)	(0.010)
	[0.020]	[0.504]	[0.075]	[0.000]	[0.554]	[0.009]
Career	-0.035	0.078**	-0.002	0.027***	0.003	-0.017**
	(0.040)	(0.026)	(0.007)	(0.006)	(0.013)	(0.009)
	[0.387]	[0.003]	[0.739]	[0.000]	[0.830]	[0.045]
Math ability	0.553***	0.469***	0.108***	0.073***	0.079***	0.002
	(0.046)	(0.044)	(0.009)	(0.010)	(0.012)	(0.011)
Verbal ability	-0.119**	-0.093**	-0.027**	-0.041***	0.076***	0.003
	(0.046)	(0.041)	(0.009)	(0.009)	(0.015)	(0.013)
Parental expectations	0.561***	0.521***	0.056***	0.034***	0.006	0.014
	(0.030)	(0.026)	(0.006)	(0.005)	(0.012)	(0.011)
Father class I and II	0.498***	0.571***	0.103***	0.095***	-0.020	-0.009
	(0.085)	(0.082)	(0.017)	(0.017)	(0.015)	(0.014)
Mother's age at birth	-0.009	0.047**	-0.001	0.011**	-0.012	-0.010
	(0.025)	(0.022)	(0.005)	(0.005)	(0.008)	(0.008)
Living conditions	-0.012	-0.026	0.001	0.003	-0.005	0.004
	(0.026)	(0.020)	(0.004)	(0.004)	(0.012)	(0.010)
Constant	16.943***	16.850***	0.136***	0.117***	6.197***	6.118***
	(0.037)	(0.031)	(0.007)	(0.007)	(0.017)	(0.015)
R-squared	0.350	0.356	0.272	0.208	0.086	0.028
Observations	4015	3895	4024	3902	621	486

OLS estimates with robust standard errors in parentheses; p-values reported in square brackets (for aspirations variables only). Degree returns are measured using median earnings across degree subjects in the 2004 APS.

Table 8: Aspirations and family outcomes

	I. Children at age 25		II. Number of children		III. Age at birth of the first child	
	Males	Females	Males	Females	Males	Females
Aspiration theme weights (base category - Family):						
Hobbies	0.012	-0.030**	-0.016	-0.058**	0.013	0.443***
	(0.010)	(0.010)	(0.032)	(0.025)	(0.156)	(0.123)
	[0.261]	[0.003]	[0.613]	[0.023]	[0.933]	[0.000]
Money	0.018	-0.023**	-0.021	-0.044*	-0.081	0.029
	(0.011)	(0.010)	(0.036)	(0.027)	(0.168)	(0.118)
	[0.105]	[0.023]	[0.556]	[0.098]	[0.632]	[0.803]
Career	0.014	-0.014	-0.016	-0.013	-0.092	0.106
	(0.011)	(0.009)	(0.035)	(0.024)	(0.167)	(0.109)
	[0.209]	[0.126]	[0.657]	[0.592]	[0.582]	[0.329]
Math ability	-0.001	-0.034**	0.036	0.006	0.246	0.537***
	(0.011)	(0.013)	(0.036)	(0.036)	(0.174)	(0.162)
Verbal ability	0.051***	0.057***	0.031	0.104**	-0.763***	-0.651***
	(0.012)	(0.014)	(0.038)	(0.039)	(0.190)	(0.165)
Parental expectations	-0.060***	-0.079***	-0.053*	-0.081**	0.748***	0.939***
	(0.009)	(0.010)	(0.028)	(0.025)	(0.133)	(0.111)
Father class I and II	-0.051**	-0.069***	0.059	0.087	0.584**	0.830**
	(0.017)	(0.020)	(0.058)	(0.057)	(0.276)	(0.256)
Mother's age at birth	-0.005	-0.050***	-0.058**	-0.087***	0.098	0.512***
	(0.007)	(0.008)	(0.021)	(0.020)	(0.104)	(0.094)
Living conditions	0.019**	0.017**	0.011	0.022	-0.442***	-0.313***
	(0.008)	(0.008)	(0.024)	(0.021)	(0.117)	(0.091)
Constant	0.279***	0.453***	1.659***	1.730***	27.582***	25.057***
	(0.010)	(0.010)	(0.030)	(0.027)	(0.148)	(0.122)
R-squared	0.066	0.130	0.004	0.024	0.100	0.175
Observations	4045	3911	4032	3905	2855	3049

OLS estimates with robust standard errors in parentheses; p-values reported in square brackets (for aspirations variables only). The dependent variables are specified as a binary indicator for having any biological children by age 25; the total number of children by age 50; and the cohort member's age at the birth of their first biological child.

Table 9: Aspirations and employment outcomes

	I. Weekly earnings		II. Hours worked		III. Number of jobs	
	Males	Females	Males	Females	Males	Females
Aspiration theme weights (base category - Family):						
Hobbies	-1.926	-0.653	0.284	0.419*	0.001	0.136**
	(1.567)	(1.032)	(0.295)	(0.222)	(0.049)	(0.043)
	[0.219]	[0.527]	[0.336]	[0.059]	[0.980]	[0.002]
Money	1.059	2.214**	0.592*	0.311	0.033	-0.001
	(1.750)	(1.005)	(0.303)	(0.281)	(0.054)	(0.041)
	[0.545]	[0.028]	[0.051]	[0.269]	[0.540]	[0.975]
Career	-0.918	1.516*	0.489	0.050	0.099*	0.006
	(1.690)	(0.899)	(0.318)	(0.225)	(0.054)	(0.036)
	[0.587]	[0.092]	[0.125]	[0.823]	[0.068]	[0.864]
Math ability	1.126	7.554***	-0.451	0.735**	-0.455***	-0.379***
	(1.736)	(1.395)	(0.340)	(0.288)	(0.053)	(0.054)
Verbal ability	-5.642**	-5.713***	0.209	0.116	-0.097*	-0.227***
	(1.829)	(1.402)	(0.339)	(0.354)	(0.058)	(0.060)
Parental expectations	-1.911	6.493***	-1.058***	0.012	-0.244***	-0.075**
	(1.274)	(0.895)	(0.244)	(0.243)	(0.042)	(0.037)
Father class I and II	-16.074***	-1.559	1.298**	0.518	0.117	0.103
	(2.776)	(2.245)	(0.589)	(0.506)	(0.086)	(0.087)
Mother's age at birth	-0.699	2.319**	-0.533**	0.489**	-0.045	-0.068**
	(1.088)	(0.732)	(0.182)	(0.177)	(0.034)	(0.029)
Living conditions	-2.076	-2.471***	0.407*	0.133	0.082**	0.071**
	(1.263)	(0.722)	(0.244)	(0.239)	(0.039)	(0.032)
Constant	91.816***	51.615***	43.155***	36.370***	2.869***	2.673***
	(1.454)	(1.050)	(0.266)	(0.262)	(0.045)	(0.041)
R-squared	0.019	0.135	0.025	0.013	0.085	0.027
Observations	3428	3460	2708	2205	3428	3460

OLS estimates with robust standard errors in parentheses; p-values reported in square brackets (for aspirations variables only). The dependent variables, all measured at age 23 (Sweep 4), are gross weekly earnings from all jobs; the weekly hours of work at the main job; and the number of job held by the time of Sweep 4 interview.

Table 10: Aspirations, savings and assets

	I. Savings and investments		II. Home ownership		III. Home value	
	Males	Females	Males	Females	Males	Females
Aspiration theme weights (base category - Family):						
Hobbies	178.242*	-107.446*	0.006	-0.031**	0.034	-0.315**
	(96.517)	(60.772)	(0.010)	(0.011)	(0.103)	(0.113)
	[0.065]	[0.077]	[0.529]	[0.005]	[0.738]	[0.005]
Money	245.041**	106.986	0.024**	0.001	0.242**	-0.028
	(108.852)	(103.935)	(0.011)	(0.011)	(0.115)	(0.114)
	[0.024]	[0.303]	[0.033]	[0.939]	[0.036]	[0.809]
Career	165.587	54.136	0.015	-0.003	0.143	-0.054
	(103.627)	(87.422)	(0.011)	(0.010)	(0.115)	(0.102)
	[0.110]	[0.536]	[0.184]	[0.743]	[0.210]	[0.594]
Math ability	204.745**	-22.176	0.036**	0.016	0.328**	0.193
	(104.259)	(133.284)	(0.012)	(0.014)	(0.117)	(0.150)
Verbal ability	203.466	-50.629	0.004	-0.021	0.018	-0.236
	(128.186)	(156.051)	(0.012)	(0.015)	(0.121)	(0.157)
Parental expectations	138.054	237.007**	-0.031***	0.005	-0.249**	0.051
	(93.092)	(72.751)	(0.009)	(0.010)	(0.089)	(0.104)
Father class I and II	869.536***	655.317**	-0.037**	-0.045*	-0.492**	-0.446*
	(254.722)	(256.468)	(0.018)	(0.023)	(0.187)	(0.240)
Mother's age at birth	-35.312	2.273	-0.004	-0.015*	-0.042	-0.190**
	(69.327)	(73.999)	(0.007)	(0.008)	(0.073)	(0.085)
Living conditions	-174.417***	-151.580**	-0.031***	-0.053***	-0.328***	-0.525***
	(51.805)	(69.135)	(0.007)	(0.008)	(0.073)	(0.082)
Constant	822.232***	955.515***	0.207***	0.364***	2.079***	3.648***
	(73.861)	(88.884)	(0.010)	(0.011)	(0.100)	(0.114)
R-squared	0.018	0.011	0.013	0.021	0.014	0.022
Observations	3428	3460	3428	3459	3428	3460

OLS estimates with robust standard errors in parentheses; p-values reported in square brackets (for aspirations variables only). The dependent variables, all measured at age 23 (Sweep 4), are specified as the total value of savings and investments; a binary indicator indicator for home ownership; and the value of primary residence in 1981 prices.