NBER WORKING PAPER SERIES

AMERICAN LIFE HISTORIES

David Lagakos
Stelios Michalopoulos
Hans-Joachim Voth

Working Paper 33373
http://www.nber.org/papers/w33373

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2025

For helpful comments, we thank Ran Abramitzky, Abhishek Arora, Elliott Ash, Leah Boustan, Leo Bursztyn, Martin Fiszbein, Vicky Fouka, Ed Glaeser, Grant Goehring, Claudia Goldin, Stefan Heblich, Pat Kline, Bob Margo, Sendil Mullainathan, Hong Wang, David Yanagizawa-Drott, Noam Yuchtman and seminar/conference participants at BU, Brown, Chicago, CUHK Shenzhen, Harvard, Oxford, Munich, the NBER Summer Institute, Stanford, Warwick, and ETH-Zurich. For outstanding research assistance, we thank Lydia Barski, Augustus Bayard, Zoe Blocher-Rubin, Catrina Brown, Vikram Dixit, Shreya Dutt, Ella Edmond, Soala Ekine, Giacomo Gattorno, Daniele Goffi, Mashod Harrison, Annie Herring, Justin Johnson, Anjali Kini, Dahlia Levine, Shraddha Mandi, Isaac Otoo, Hanna Polyak, Simone Seiner, Elijah Sumernikov, Mafo Tekobo, and Hans Xu. All potential errors are our own. We have no relevant financial relationship to disclose. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

American Life Histories
David Lagakos, Stelios Michalopoulos, and Hans-Joachim Voth
NBER Working Paper No. 33373
January 2025
JEL No. I31, N0, O10, P00, Z10

ABSTRACT

What does it take to live a meaningful life? We exploit a unique corpus of over 1,400 life narratives of older Americans collected by a team of writers during the 1930s. We combine detailed human readings with large language models (LLMs) to extract systematic information on critical junctures, sources of meaning, and overall life satisfaction. Under specific conditions, LLMs can provide responses to complex questions that are indistinguishable from those of human readers, effectively passing a version of the Turing Test. We find that sources of life meaning are more varied than previous research suggested, underlining the importance of work and community contributions in addition to family and close relationships (emphasized by earlier work). The narratives also highlight gendered disparities, with women disproportionately citing adverse family events, such as the loss of a parent, underscoring their role as keepers of the kin. Our research expands our understanding of human flourishing during a transformative period in American history and establishes a robust and scalable framework for exploring subjective well-being across diverse historical and cultural contexts.

David Lagakos
Department of Economics
Boston University
270 Bay State Road
Boston, MA 02215
and NBER
lagakos@bu.edu

Hans-Joachim Voth
Department of Economics
U of Zurich
Schoeneberggasse 1
CH-8001, Zurich
Switzerland
joachimvoth@gmail.com

Stelios Michalopoulos
Brown University
Department of Economics
64 Waterman Street
Providence, RI 02912
and CEPR
and also NBER
smichalo@brown.edu

"development is ultimately the expansion of the capacity of the population to engage in freely chosen and valued activities..."

Amartya Sen

1 Introduction

Suppose for a moment that life had an "exit interview" in which each of us got a chance, during our final years, to record a brief life history for posterity. Which topics would we choose to focus on? Would our narratives be valuable in some way for future generations? Would our collective responses help inform mankind about what brings satisfaction or meaning to life? Would descriptions of our trials and travails provide useful insights into how life experiences shape our attitudes and determine our well-being?

These questions are less fanciful than they may first appear. In the late 1930s, as part of the New Deal, the Roosevelt administration hired a team of unemployed writers to record the life histories of several thousand older Americans. The goal was to interview representatives of older generations from all backgrounds and to "let them tell their own stories" (Harrington, Kerr and Alsberg, 1939). The writers received minimal guidance on the structure of the interview and had no scripts. Collectively, the narratives were meant to be "history from the bottom up" and to "hold up a mirror to America" (Banks, 1980). The result was the *American Life Histories*, which went on to become a seminal output of the New Deal's Federal Writer's Project. Nothing quite like it was produced before or since.

In this paper, we provide the first systematic analysis of the American Life Histories. The paper has two broad goals. The first is to develop new methods that allow for systematic, comparable, and replicable classification of human sentiments in complex narratives. What sets narratives such as these apart is the way in which they capture the deeper sense of meaning and purpose that individuals derive from their experiences. Yet there are no off-the-shelf ways of extracting this type of "higher-level" information from documents like these. Nor can one simply rely uncritically on large language models (LLMs), as they are known to hallucinate, introduce biases, or otherwise misinterpret complex text inputs (see e.g. Dell, Forthcoming; Korinek, 2023). The emerging literature in computer science emphasizes that the analysis of emotions and motivations in text is challenging even for frontier LLMs, which tend to underperform humans (Amirizaniani et al.,

2024). As we describe below, our methods allow us to succeed at this challenge, combining the latest techniques in LLMs and multistage assessments with extensive human readings in order to extract information reliably at scale.

The second goal of the paper is to use our methods to answer questions about the sources of life satisfaction and the role of life experiences. The American Life Histories offer a unique opportunity to examine such issues based on a large number of comparable recollections, collected in the aftermath of the Great Depression. As Hirsch (2003) puts it, in his history of the Federal Writer's Project: "A major value of these life histories is that they reveal the shape people gave to their experience and thus the meaning they found in their lives." How to measure well-being is a core question in economics. Dissatisfaction with the limitations inherent in GDP has led to a plethora of alternative measures designed to capture the determinants of human 'flourishing,' with the U.N.'s Human Development Index (HDI) as one prominent example (Fitoussi, Stiglitz and Sen, 2011). We contribute by compiling systematic information on what brought satisfaction, happiness, or meaning to the lives of our narrators. We then record the critical junctures, or key turning points, in their lives, and gauge whether these events were primarily positive or negative. Finally, we score the overall retrospective life satisfaction of each respondent, using a cardinal metric. For each outcome of interest, we report the average pattern and offer a decomposition along gender, race, occupational status, urban or rural location, immigration status, and internal migration status.

To fix ideas, consider the lives of two narrators in the American Life Histories corpus. Minnie Marshall was born in Virginia to a single mother. Her life history is a Dickensian account of resilience in the face of adversity. After losing her mother at a young age, Minnie moves to New York in search of work. Her frail body leaves her ill-suited for long hours of manual labor, resulting in her repeatedly being fired. An early marriage ends badly after her husband abuses her violently, knocking out her front teeth. Minnie perseveres, but struggles to make ends meet working as a day laborer and domestic servant. Wages are low, Minnie acknowledges, "but whut ah'm gonna do? Ah got to live— got to hab a place to steh."

Elmo Acosta was born in Jacksonville, Florida and began work as newspaper boy at age six. He eventually becomes the owner of a grocery business and gets "stung by the political bee," entering local politics. Most of his life history proudly details his contributions as a member of the City Council and as Parks Commissioner, including swamp clearings, city beautification projects, founding the city's nursery, and much else. "The biggest thing in my life?" he concludes: "Making it possible [to build] the great St. John River bridge for the people of Duval County."

To record and analyze the rich content of these narratives, we develop and apply methods to combine selected human readings with the power of LLMs, extracting higher-level text features in an objective, replicable, and transparent manner. We first create a human audit sample of 16,200 hu-

man observations about representative higher-level characteristics of the narrators that we recorded as answers to yes-no questions, such as whether they showed empathy toward others. We then illustrate the circumstances under which LLM answers are indistinguishable from human answers using a version of the Turing (1950) test. We show that LLM answers only become indistinguishable from human answers when the text is broken into sufficiently small "chunks." In contrast, LLM answers about the entire document, and traditional dictionary methods, fail.

We then turn to more open-ended questions, first focusing on identifying the sources of happiness, satisfaction, meaning, or purpose in the lives of each narrator. We show that these open-ended questions are significantly more challenging for LLMs; answers typically fail the Turing Test. Our preferred method uses a form of multistage assessment, combining the output from multiple calls to LLMs.[1] The process works as follows: First, we ask an open-ended question to one LLM, breaking the text into chunks as before. We also ask the LLM to provide a numerical confidence score on its own answer. Next, we call on a second LLM to read the output of the first, examine whether the cited text supports the first LLM's assessment, and then filter out low-quality answers. We also filter out answers that fall below a pre-determined confidence threshold. Using this method, LLM answers about the sources of meaning in life pass the Turing Test when compared to answers from four human readers in 140 life histories: the LLM answer is as likely to agree or disagree with the majority human assessment as any human RA.[2]

When scaling up to the full set of life histories, we find some sources of happiness and meaning familiar to the literature, such as the joys of family life, or the bonds offered by social connections with friends and the community, as emphasized in the famous Harvard Study of Adult Development (Vaillant, 2012, 2008). Other themes are less familiar, such as the serenity of being in nature or the adventures lived through when younger. Surprisingly, around half of the narrators emphasized the meaning they found in their work – whether through the pride in a job well done, the skills they developed, or the recognition by bosses and peers. In contrast to the conclusions of the Harvard studies, work figures as prominently in the life narratives as family and community – if not more so – and is present in similar measure for men and women, and along racial and regional lines.

Sources of satisfaction and meaning vary across sub-groups of the population in notable ways. Women were more likely than men to find meaning in religion and resilience after difficult circumstances. Those with occupations higher up the socioeconomic ladder found more purpose in their work, suggesting that "better" jobs offer not only higher pay but more non-financial rewards.

[1] This is a simplified version of what is known as "agentic workflow" (Singh et al., 2024).

[2] Note that our approach does not require us to take the information in the life histories at face value; by focusing on the subjective perspective of the respondents we side-step one of the key pitfalls of oral history (Shopes, 2002).

Individuals migrating within the United States were more likely to describe satisfaction from educational opportunities, particularly among those migrating to cities. Sources of meaning also echo the traditional folklore of their ancestors (as recorded by Michalopoulos and Xue, 2021). For instance, narrators whose oral traditions are rich with tales and legends depicting successful resolution of challenges are more likely to recount resilience-related episodes in their life histories. Similarly, gender gaps in having found meaning from work are stronger among narrators tracing their ancestry to societies where males and females occupy traditional gender roles in the oral tradition.

A common thread in the narrators' recounting of their lives is the focus on positive or negative turning points or structural breaks. It appears that these pivotal events allowed our narrators to reconstruct their life journeys. Discussions of critical junctures in life – such as migration episodes or health shocks – made up a significant portion of most narratives. We find that career, migration, and community junctures were generally positive, while those related to family, historical events, health, and nature were predominantly negative. The life histories of women were generally bleaker than those of men, with family-related breaking points, in particular, being more often negative than positive. This is consistent with women being the 'keepers of their kin,' and more likely to bear the burden of adverse shocks, including the loss of a parent when young. Migrants showed generally brighter outlooks than non-migrants, with a significantly higher proportion of positive critical junctures related to work, education, and family. City dwellers experienced more negative work-related critical junctures than rural folk, highlighting the risks of wage work versus family farming during this era.

We measure retrospective life satisfaction among our narrators by asking the LLM and a set of human readers to score how happy each narrator's life seemed to be on a scale of 1 to 7. After showing that the LLM scores match well with the human ones, we look at how happiness levels differ across groups. Those who worked in occupations further up the socioeconomic ladder scored higher on average, consistent with a large economic literature measuring contemporaneous life satisfaction. Women and men had similar scores overall, and white narrators had higher average scores than non-whites, consistent with their better work outcomes and greater freedoms in this era. Those finding meaning in their jobs were happier looking back than those indicating that hard work is important, highlighting the subtleties of stances toward one's labors conveyed by our narrators. In terms of magnitudes, the sources of life satisfaction and critical junctures are not all equally important predictors of overall happiness, in a statistical sense. Having found meaning in one's work and the community are the two strongest predictors of overall happiness; negative financial, health, and community shocks translate systematically into lower scores.

Finally, to describe the joint patterns of critical junctures and sources of life meaning, we classify

the narratives into distinct groups with common properties using agglomerative hierarchical clustering methods. We show that the narratives can be parsimoniously described using six clusters, four generally positive and two more negative. The most common sources of meaning for the positive clusters are community; work; family and adventures; and education, respectively. The two negative ones offer few signs of life satisfaction or meaning, and are most likely to highlight negative critical junctures. We show that the clusters divide the narratives along gender and racial lines, along with urban-rural, immigration, and internal migration status. Among the positive groups, migrants are disproportionately found in the one focused on education. Among the two negative groups, women are far more concentrated in the one focused more on negative shocks linked to family or career.

Related Literature Our work relates to a larger literature using text-to-data methods (Ash and Hansen, 2023; Gentzkow, Kelly and Taddy, 2019; Beach and Hanlon, 2023).[3] It contributes to the small but growing body of work that uses LLMs to create and analyze economic data (see Korinek, 2023, and the references therein). LLMs are easy to use, but lack transparency and can perform worse than alternative text classification methods using domain-specific training (see Dell, Forthcoming, Table 2).[4]

We also make a contribution relative to the emerging computer science literature using LLMs to reason about human emotions and motivations, which requires a "Theory of Mind" or ToM (Zhou et al., 2022; Ma et al., 2023; Strachan et al., 2024). Most standard benchmark tests used by computer scientists to assess the performance of LLMs focus on questions with a clear right or wrong answer. Even the most advanced LLMs appear to struggle with nuanced assessments and open-ended questions, especially when ToM is required (Amirizaniani et al., 2024). Our methods show the potential for solving this challenge through a judicious use of multi-stage assessment and extensive human validation. As a result, LLM answers can become indistinguishable from human answers.

Our work, and in particular our analysis of Yes-No questions in Section 3, takes a similar approach to Bartik, Gupta and Milo (2024) who use an LLM to record features of building codes across U.S. counties. Adukia and Harrison (2024) use methods related to ours to help record the topics present in textbooks. Geiecke and Jaravel (2024) show that LLMs can be used effectively to interview respondents, including about complex topics. To our knowledge, our systematic comparisons of

[3]Prominent examples include work on economic uncertainty Baker, Bloom and Davis (2016), corporate earnings calls Hassan et al. (Forthcoming), patents (Kalyani et al., Forthcoming; Kalyani, 2024), ancestral folkore (Michalopoulos and Xue, 2021), and monetary policy statements (Handlan, 2024), to mention but a few.

[4]For example, RoBERTa outperforms GPT-4 in answering Yes-No questions about the content of news articles in the digital database of historical U.S. newspapers constructed by Dell et al. (2023).

LLM to human output using a version of the Turing Test and our use of agentic workflow are novel to the economics literature and broadly consistent with the intuitive guidelines suggested by Ludwig, Mullainathan and Rambachan (2025).

The retrospective nature of the narratives we analyze connects to the burgeoning literature in memory studies. What and how an individual remembers is at the heart of studies in cognitive psychology; see Kahana (2012), Bordalo, Gennaioli and Shleifer (2020), and Bordalo et al. (Forthcoming) for applications in behavioral economics. Our paper also builds on recent work in economics that collects systematic data on subjective beliefs, views, and reasoning (Akerlof and Kranton, 2000; Besley and Persson, 2023; Enke, 2024; Stantcheva, 2021), and uses textual information in historical documents, such as names, to shed light on individual attitudes (Bazzi, Fiszbein and Gebresilasse, 2020; Andersen and Bentzen, 2022; Knudsen, 2019). In this way, we add to work attempting to measure and understand the broader determinants of well-being among workers and individuals (e.g. Di Tella and MacCulloch, 2006; Stevenson and Wolfers, 2008; Frey, 2010; Jones and Klenow, 2016). Our finding that work is a major source of meaning and purpose supports the idea that intrinsic rewards from working may be more important than economists typically think (Ariely, Kamenica and Prelec, 2008; Besley and Ghatak, 2005; Cassar and Meier, 2018; Kaplan and Schulhofer-Wohl, 2018; Maideu-Morera, 2024). Dal Bó, Finan and Rossi (2013) and Ashraf, Bandiera and Jack (2014) demonstrate that intrinsic motivation is important for many public servants, and pro-social goals more generally are associated with more purpose at work (Cassar and Meier, 2018). Ashraf et al. (2024) study meaning at work using personal statements, similar in some ways to our life narratives.

Our work is related to, but distinct from, the literature on economics focusing on narratives as rationales for decisions or observed phenomena like economic downturns or inflation episodes (Shiller, 2017; Benabou, Falk and Tirole, 2020; Flynn and Sastry, 2024; Andre et al., 2024). It also differs from the psychology literature on "life narratives" as determinants of identity (see McAdams, 2001; Ayyar et al., 2024). Our research contributes to a growing literature that uses the text of historical documents to shed light on contemporary outcomes (Michalopoulos and Rauh, 2024; Almelhem et al., 2023; Michalopoulos and Xue, 2021). Recent advances in machine learning and the digitization of hitherto inaccessible materials open up the possibility of analyzing unstructured oral histories that have remained beyond the reach of social scientists, using open-ended questions for inductive hypothesis generation (Ludwig and Mullainathan, 2024; Zhong et al., 2023). We hope that our methodological contribution will enable researchers to rely less on secondary sources, the evaluations of famous scholars, or texts created by elite members, and help to give voice to the personal recollections of ordinary men and women.

2 The Federal Writers Project's *American Life Histories*

We begin by describing the life histories that are the subject of our paper and summarizing some of their basic characteristics.

2.1 The American Life Histories

The Federal Writers Project was created during the New Deal to generate work for the unemployed. It hired around 6,500 people at its peak, making it a relatively modest employment program compared to the 8 million people employed by the Works Progress Administration. Still, the Federal Writers Project was described the "biggest literary project in history" (Time Magazine). The main goal of the project was what they wrote was considered a secondary objective. Its mandate required that 90 percent of its workers be hired from the unemployed ranks (Mangione, 1996; Borchert, 2021). As a result, most writers were as ordinary as the people they interviewed.

The Federal Writers produced several outputs of note. The largest and most widely known was the *American Guide Series*, which consisted of illustrated travel guides to each of the fifty states (see Hirsch, 2003; Taylor, 2009; Griswold, 2016; Borchert, 2021). The *Slave Narrative Collection* was a collection of interviews with former enslaved peoples (see Botkin, 1945; Escott, 1979; Crawford, 1980). One pioneering output of the Federal Writer's Project, and the focus of this project, was the *American Life Histories*, which recorded narratives about the lives of a broad set of U.S. residents, most of them older and looking back at their lives. The writers hired for this task were given little guidance as to what to ask or record about the subjects (Harrington, Kerr and Alsberg, 1939). As director W.T. Crouch put it when describing the project, "with all our talk about democracy, it seems not inappropriate to let the people speak for themselves..." The resulting interviews were the first of their kind.[5]

The objective was to interview a broad and representative set of Americans. An instructional manual for the writers specified that they should interview "tenant farmers, farm owners and their families, cotton mill villagers and their families, persons and their families in service occupations in towns and cities, and persons and their families in miscellaneous occupations such as mining, fishing, lumbering, and turpentining" (Harrington, Kerr and Alsberg, 1939). The goal was "readable and faithful representations of living persons, which, taken together, will give a fair picture of the structure and working of society." In one correspondence with a writer interviewing mill

[5]The *American Life Histories* differ from other oral histories from this era, such as Studs Terkel's oral histories of the Great Depression (Terkel, 1970), which focus largely on how people coped during the years of the depression rather than recording a history of their lives.

workers, project director W.T. Couch wrote that "we are not trying to prove anything, but on the contrary are trying to get an honest and accurate picture of mill life as possible." (Hirsch, 2003).

Less is known about how respondents were selected in practice, and remaining accounts point to a mix of personal networks and chance encounters. Banks (1980, p. xvii) writes that "Federal Writers were encouraged to arrange interviews through community or work-related organization, but that suggestion was often ignored in favor of chance contacts." Some writers seemed to have interviewed their relatives, and many attempted to interview strangers in their local communities. One wrote that "I hung around playgrounds; I hung around the street, the bars. I went into hundreds of apartment buildings and just knocked on doors." According to Hirsch (2003, p. 171), "Field workers approached those they did not know in a casual and random manner." Just how representative their sample was, relative to the U.S. adult population, is a question we take up in the following section.

The *American Life Histories* have been digitized by the Library of Congress and are freely available via their website. There are 2,900 individual documents available for download. However, many of these documents are not actually life histories, but rather descriptions of places, events, or other miscellaneous folklore. Since the focus of this project is the lives of individual people, our first step was to read each document to determine which were life histories and which were not. We did this by hand, ending up with a total of 1,799 documents that we classified as life histories. Many of these turned out to be duplicates or similar versions of the same life history. We manually merged all documents that covered the same narrator giving a total 1,478 unique life narratives, each focusing on the life of one individual. In our analysis, we restrict attention to the 1,335 life narratives about an adult over the age of 18 that have between 500 and 10,000 words. The vast majority of these narratives were recorded in 1938 or 1939.

2.2 Three Lives Illustrated

To illustrate the richness of these documents, we summarize three life histories in detail. As a frame of reference, we include information about the narrators available in the 1930 census.

Minnie Marshall. We begin with Minnie Marshall, who, according to her census record from 1930, is a 34-year-old single black female living in rented housing on West 131st Street in Manhattan, and working as a domestic servant.

Minnie's interview was recorded in the Bronx in 1938. She was born in Virginia, and we learn that her childhood was a difficult one. She never knew her father. Her mother describes him as a sailor who "had gone down to sea with his ship." Minnie's mother works long hours to keep the

family afloat, and Minnie helps out with the work while sporadically attending school. Tragedy hits when Minnie's mother takes ill and Minnie is "yanked out of school in the third grade at the age of fourteen" in order to take her "ailing mother's job at Miss Sarah's." Minnie's mother dies a few days later.

As a frail teenager, Minnie moves to New York City in search of better job opportunities. She finds jobs but struggles to keep them, repeatedly being "fired from various positions due to lack of experience and youth — not having enough endurance and muscle for fifteen to eighteen hours of strenuous laundry and housework." Minnie tries her hand at marriage at the age of 16. It does not go well. Minnie painfully describes the abuse she received at the hands of her spouse, and the breakup that occurs when her husband attacks her, "knocking out her front teeth, and taking all of the money, stumbled, and disappeared into the night." She never sees him again.

A resilient woman, to say the least, Minnie continues to work long hours as a domestic servant to make ends meet. She tries to find better paid work through an employment agency, but the agency demands payments for jobs, which Minnie cannot afford. She continues to search for jobs, taking anything she can find. Wages are barely enough to live on, but, as Minnie puts it: "whut ah'm gonna do? Ah got to live— got to hab a place to steh."

Minnie is 34 at the time of her interview but "looked 45." One can see why. Her heart-breaking narrative illustrates the travails and dearth of opportunities in her life.

Elmo Acosta. According to the 1930 census, Elmo Acosta is a 66-year old white male living in Jacksonville, FL, and working as a grocer. He is interviewed in 1938 by the Federal Writer's Project. Elmo also seems to have had a difficult childhood. His mother dies when he is two, and his father when Elmo is seventeen. Elmo starts work at age six, selling papers, though he does manage eventually to attend secondary school.

Elmo works at various jobs in his early career, including as a railroad clerk. After getting passed over for a promotion there twice, he quits and opens his own wholesale grocery business. Eventually he gets "stung by the 'political bee," and enters the world of local politics. Most of Elmo's narrative describes his contributions and accomplishments through public service. Elmo's political career in Jacksonville – then a city of 30,000 people – begins when he is elected to city council. He later joins the city commission, and finally becomes Parks Commission for the city.

Elmo proudly recounts how he "advocated the planting of holly, oak, and magnolia trees as a memorial to the soldiers of the World War. . . ." He also directed the filling in of swamps near Jacksonville and established the city nursery, which Elmo boasts as "the only one in the United States, by the way." He enlarged the city's park lands to a total of 300 acres during his tenure.

Elmo mentions his children with pride, several of whom graduated from the University of Florida. But he concludes his narrative by returning to his public service: "the biggest thing in my life? Making it possible [to build] the great St. John River bridge for the people of Duval County."

Jeff Waggoner. Moving west, we turn to the life of Jeff Waggoner. In the 1930 census, Jeff was a literate 78 year old, retired, and married to Mary Waggoner. He was born in Texas and lived in Wichita Falls, TX, on a property worth $20,000 (equivalent to $ 310,000 in 2020 US dollars adjusting for inflation, or $1,720,000 as a share of national wealth). Jeff is interviewed by the Federal Writer's Project in 1938 at his home in Wichita Falls. The opening line of his narrative reveals a lot about his past: "When I think of those days as a cowboy on the range I get real lonesome."

Jeff takes his first cowboy job at age 18 and spends seven years riding the plains and sleeping in dugouts during the winters. His company would send provisions twice a winter and otherwise Jeff and his fellow cowboys would kill beef as needed to survive until spring. The fact that he spent the rest of his career as a cowboy shows how suited he was for this line of work.

Most of Jeff's narrative describes his life as a cowboy. During the daytime, he would ride all day with his team, leading the cattle north from Texas. At night, they "sat around the camp fire and listened to one of them playing the fiddle. Sometimes they all sang the cowboy songs." Jeff explains that singing songs in the evenings was a form of comradery between the cowboys, but also served the valuable purpose of calming the cattle. Stampedes were both dangerous and bad for business: cows gain more weight walking and eating grass then they do stampeding.

Jeff's narrative reveals that he got satisfaction from cowboy work itself. As he put it:

> I enjoyed the trips to Kansas City. There were eight cowboys, a boss, a cook and a horse wrangler in the expedition. The boss traveled ahead and located water... We boys drove the cattle till they got to the wagon. We made about twenty miles a day.

He also enjoyed the occasional adventure that came with the cowboy lifestyle, including tracking down cattle that had gone off stampeding and bringing them back on track.

Jeff married Mary Cose of Decatur on February 23, 1888, describing it as "about the wisest thing I ever did." Eventually, he acquires his own stock which accumulated over time, and his own large ranch, leaving Jeff and his family in a relatively comfortable position by the time of the interview.

2.3 Basic Characteristics of the Narrators

We recorded the basic characteristics of each narrator using a mix of human and computer readings. We began by hand recording the name, sex, age (often approximate), birthplace, and interview location of each of the life narratives. We attempted to record the race and ancestry of each narrator, which is often revealed indirectly, like when a narrator mentions that their parents were Italian immigrants. For a more detailed description of our methods, see the Data Appendix.

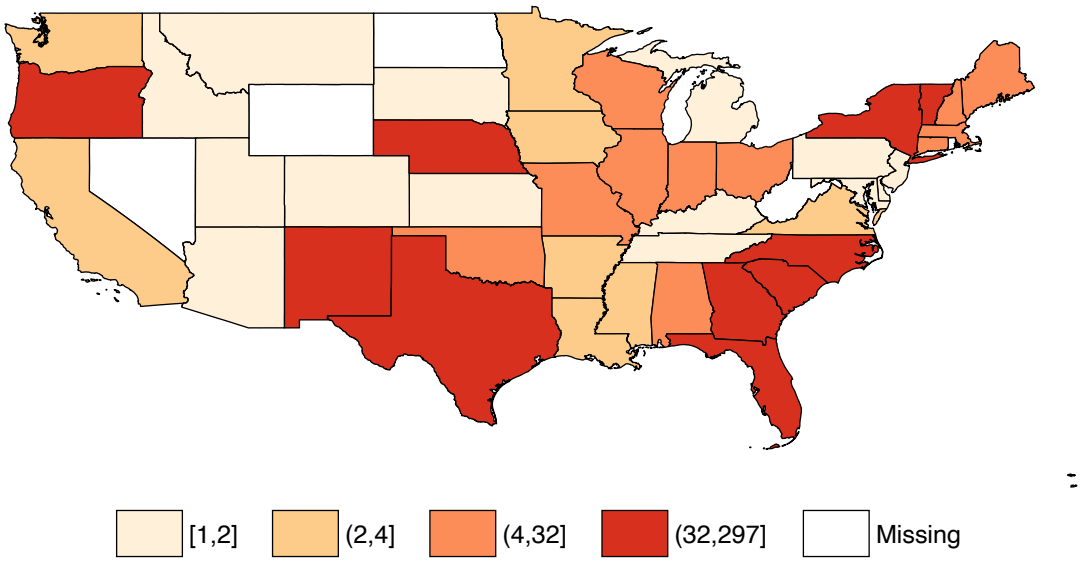
Like the residents of the United States at the time, the narrators came from a wide variety of mostly European backgrounds (see Appendix Figure A3). When an individual is born in the US, no other ancestry is mentioned, and their last name is of English descent, we assign them to American ancestry, which is the numerically dominant one. Significant numbers are also of German, Italian, Irish, Swedish, Greek, and French descent. A total of 34 ancestries were found.

Many narrators described working in multiple occupations throughout their lives. We used GPT-4 to help record all occupations for each narrator during both their early years as well as during their prime working years. This resulted in short but non-standardized text descriptions such as "general store owner" or "cattle rancher." We then mapped these descriptions into the 1950 census occupation categories and obtained the associated Duncan's socioeconomic index (SEI) for each occupation for each narrator. Almost 69 percent of the 1,335 individuals changed occupations during their lives, consistent with the high rates of job turnover common in early 20th-century labor markets (Fishback, 1998).

Similarly, with a mix of human readers and GPT-4's help we reconstructed the spatial trajectory of each narrator's life history, by assigning, to the best of our understanding, a location to every year of their life. We used the Google Maps API to geo-code each record location and mapped it to census counties following the 2010 Census delineation. We then divided our narrators between movers and stayers by tracking each individual across years at the county or state level. We define someone as a mover if we recorded at least one state or a county change, and for those who moved we further document the exact year of the first, second, and third move. In total, out of the 1,335 narrators, 703 moved within the United States during their lives reflecting the substantial internal migration of the late 19th century and early 20th century.

Figure 1 displays the locations of 1,335 narrators during their productive years. For 39 of the 48 states, we have at least one narrator who spent her productive years there, distributed across the Northeast (16 percent), Midwest (15 percent), South (32 percent), Southwest (32 percent), and West (5 percent). Compared to the 1940 U.S. census population distribution, narrators were more concentrated in the South. Texas, New Mexico, Nebraska, South Carolina, North Carolina, New York, Georgia, and Florida provided the largest number of life histories.

Figure 1: Location of the Narrators during their Productive Years



The writers' descriptions of the narrators' housing feature consistently in the life histories, providing crude but relatively objective indicators of prosperity at the time of interviews. Using GPT-4, we classify housing quality into three broad categories based on reported size and quality. For example, a "dilapidated shack" falls into the lowest category, a "three-room cabin" belongs in the middle one, and a "comfortable brick residence" falls into the highest category.

Table 1: Characteristics of Language and Writers

	Mean	Median	P10	P90
Word Count	2,804	2,054	689	5,331
Difficult Words	187	147	55	363
Flesch-Kincaid Grade	5.5	5.4	3.5	7.6
Colloquial Language	0.40	0	0	1
Female Writer	0.55	1	0	1
Famous Writer	0.02	0	0	1

Notes: This table reports linguistic characteristics of the narratives and their writers.

2.4 Linguistic Analysis and Writers

Table 1 provides an overview of the language and writers of the life histories. The average life history contains 2,804 words, with a median of 2,054 words, and 90 percent are under 5,331 words. Regarding readability, each text contains an average of 187 "difficult words" and is written at a fifth or sixth-grade comprehension level according to the Flesch-Kincaid measure. For comparison, U.S. presidential candidates typically speak at a seventh to ninth-grade level, with the notable exception of Donald Trump, who speaks at a fourth-grade level (Kayam, 2018).

One key feature of our corpus is that it captures spoken language more authentically than many other written sources. The interviewers often employed phonetic spellings to record not just the words, but also the pronunciation of their subjects. For example, some narrators' speech appears as "I'se sho glad" rather than "I sure am glad," or "ride a hoss" rather than "ride a horse." Minnie Marshall's narrative, as seen above, includes such phonetic transcriptions. We find that around 40 percent of the documents preserved the narrators' speech patterns in this way. As we show below, this colloquial language does not significantly impede GPT-4's comprehension of these texts relative to native-speaking human readers.

Most authors of the life histories were not professional writers. A 1938 internal survey of the Federal Writers' Project staff found that only 1.8% of authors had worked as writers before the project, 2.2% had held editorial positions, and 5% had sold stories to newspapers and magazines (Penkower, 1977, p. 73).[6] The narratives in our corpus were written by 258 distinct authors. The distribution of life histories per writer is highly skewed: about half the documents were written by a set of 31 "core writers," each completing an average of 23 life histories. Based on first names, we estimate that 55 percent of the interviews were conducted by female writers. While race is more difficult to determine from names alone, historical accounts suggest minimal non-white representation. For example, Penkower (1977) reports that just 2 percent of writers employed by the Federal Writer's Project were non-white; Griswold (2016) estimates that non-whites made up 4 percent of the American Guides writers.

The Federal Writers' Project notably employed a number of writers who later achieved literary fame, including Studs Terkel, Ralph Ellison, and Zora Neale Hurston. Narratives penned by (subsequently) famous writers could differ systematically from those written by ordinary writers. Famous writers, for instance, might have embedded more personal opinions or embellishments, transforming simple accounts of daily activities into more engaging – but potentially less factual – narratives. To address this concern, we measured writers' fame by their receipt of major awards

[6] Borchert (2021) recounts an anecdote where four staff at a FWP office sprung up at once to fix a leaking office toilet: all were out-of-work plumbers.

Table 2: Life History Narrators versus Adults in the U.S. Census

Percent	Life History Narrators	All Adults, 1940 Census	All Adults, Re-Weighted 1940 Census	Households Heads Over Age 50 Re-Weighted 1940 Census
Above 50 years old	79	29	27	100
Male	66	50	50	77
Non-white	14	9	15	14
Born outside USA	12	13	8	14
Urban	45	59	46	43

Note: This table reports the basic characteristics of the 1,335 narrators of the American Life Histories in column 1, all adults (age 18+) in the 1940 U.S. census in column 2, all adults in the 1940 U.S. census re-weighting by the narrator state of residence shares in the American Life Histories, and focusing on Household Heads over age 50 in the 1940 U.S. Census in column 4.

(National Book Award, Pulitzer Prize, or Nobel Prize in Literature) or placement on the New York Times Bestseller list. By this metric, only 11 of 253 writers (4.3 percent) and 31 of 1,335 interviews (2.3 percent) came from famous interviewers, aligning with contemporary descriptions of the staff as mostly ordinary out-of-work adults.

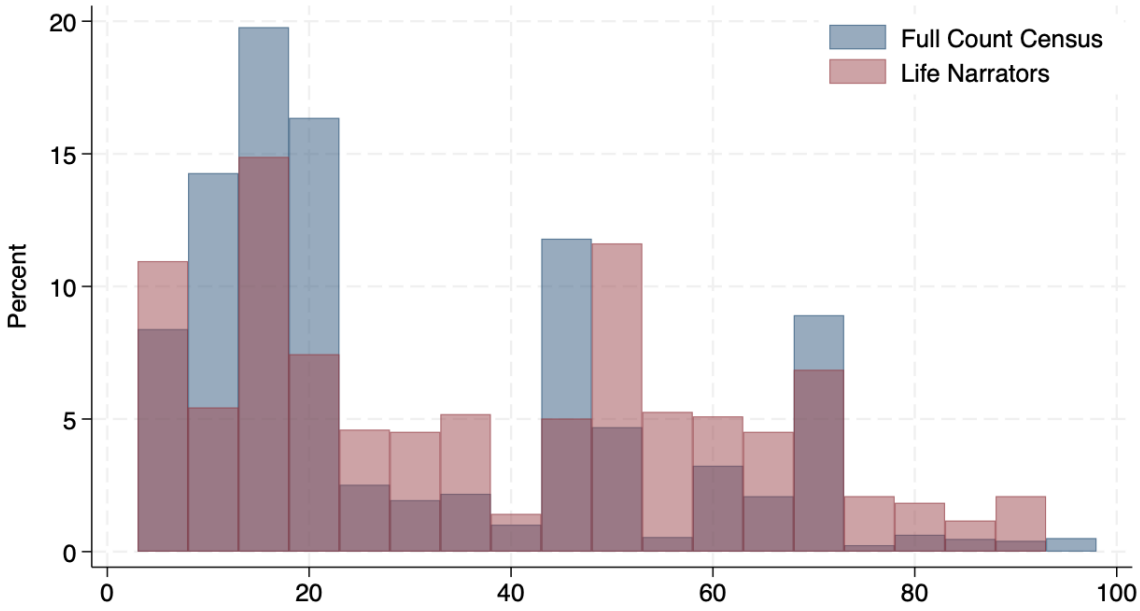
To account for potential writer influence on the narrative, we include writer fixed effects in our main analysis to follow. We also control for document length, recognizing that longer documents may reveal more information about narrators. We have explored sensitivity to excluding life histories written by famous writers but this has no material impact on any of the results below.

2.5 Comparison to the U.S. Census

We next examine how representative the narrators are compared to respondents in the U.S. census. Table 2 compares summary statistics of the 1,335 life histories (first column) with three census comparison groups (subsequent columns). About 80 percent of narrators were over 50 years old, substantially older than the general population, where only 29 percent of adults reached this age. This age distribution aligned with the project's goal of recording the experiences of older Americans whose life stories were largely complete. Appendix Figure A4 plots the complete age distribution of the narrators compared to the U.S. census.

Across all life histories, about two-thirds of the narrators were men, compared to 50 percent in the

Figure 2: Duncan Socioeconomic Indices: Narratives and U.S. Census



Note: This figure compares the Duncan Socioeconomic Index distribution of the narrators with the 1940 Census. The value for each Life Narrator was determined to align most closely with their main occupation during their prime years. The census sample includes all individuals over age 18.

census. We have not found any specific reason to suggest an intentional over-sampling of men. In some of our analyses, we assign weights so that the views of men and women are equally represented. Around 14 percent of narrators were non-white, exceeding the 9 percent among all census adults. The narrators included 12 percent foreign-born individuals, comparable to the 13 percent among all adults. Regarding location, 45 percent of narrators lived in urban centers, below the census figure of 59 percent, reflecting substantial representation from rural areas in the Southwest and Midwest, such as Nebraska.

In the third and fourth column of Table 2, we compute the census population shares when re-weighting by the state population shares in the American Life Histories. The basic idea is to ask how the narrators compare to individuals coming from the same distribution of states. We find that now the census shares align more closely with ours, with urban shares of 46 percent and non-white shares of 15 percent. The final column reports statistics for census household heads over the age of 50. These are 77 percent men, making them closer to our narrators on the gender dimension. They are also very similar in terms of race, immigrant status, and urban residence.

When comparing occupations, our narrators span the range of socio-economic outcomes recorded

in the census; averages are broady similar. Figure 2 plots the distribution of occupation scores in the two groups, measured by the Duncan Socioeconomic Index (SEI), which ranges from 1 to 100.[7] Visually, the distributions are similar in many ways, with both having significant mass below 20, which includes occupations like "rancher," "farmer" and "laborer." Neither has many scores above 60, such as "manager," "doctor," "lawyer" or "civil engineer." On average, the occupational scores are 38 among the narrators and 30 in the census. Thus, at least from an occupational perspective, the narratives do seem to "give a fair picture of the structure and working of society," which was the original goal.[8]

3 Extracting Higher-Level Info Using Yes-No Questions

This section describes our first approach to extracting "higher-level" information from the life narratives by asking yes-no questions about specific topics of interest. This approach has the advantage of being simple and transparent for human readers and, as we show below, LLMs, to answer. The main limitation is the need to preselect topics from the vast array of potential characteristics one could record. We focus on characteristics we view as important and representative, and introduce our method for comparing computer and human answers using a version of Turing's Indistinguishability Test.

3.1 Representative Higher-Level Characteristics

We list the characteristics in Table 3, alongside example passages that illustrate text indicating a positive response. The first three characteristics represent negative critical junctures in the narrators' lives, as we discuss further below. They are: the death of a parent, the death of a parent during the narrator's youth, and poverty forcing the narrator to leave school. We intentionally picked the first two to be closely related, so as to assess the ability of the large language model to distinguish the two. While the timing of a parental death may be subtle in terms of language, the economic consequences differ dramatically between loss at age 5 versus age 50. Poverty ending one's schooling is more nuanced and subjective than parental loss.

[7] High-frequency bars in the Census distribution in the histogram are driven by the following OCC1950 categories in Ruggles et al (2024): "laborers (nec)", 9.2 percent of the adult population with SEI equal to 8; "farmers (owners and tenants)", 10.5 percent with SEI equal to 14; "operative and kindred workers (nec)", 9.7 percent with a SEI of 18; "salesmen and sales clerks (nec)", 5.2 percent with a SEI of 47; and "managers, officials, and proprietors (nec)", 6.4% with a SEI of 68. For each occupation according to the 1950's classification, the highest SEI was assigned. If the recorded occupational response was "N/A", the SEI was set to 0.

[8] Appendix Figure A5 plots the occupation distribution for household heads over 50, which has similar patterns.

Table 3: Representative Higher-Level Characteristics

Characteristic	Example Passage
Parent died	"My mother died several years ago at the age of seventy-five."
Parent died when I was young	"My father died when I was nine years old, that was in 1879."
Poverty kept me out of school	"I quit school in the sixth grade to take up the canning plant work. My daddy's health failed and he lost what money he had so my eldest brother and me started to work."
Religion is important	"We always tithe, and when the Lord gives us an extra blessing we divide that also with our church."
Hard work is important	"I believe it can be proved that I've sold more cattle than any other cattleman in boots."
Agreeable personality	"It's true we razzed on each other a lot but I never saw a cowboy that wouldn't stand up for another when he needed help or was in any kind of a jam.
Shows empathy towards others	"We never had any children of our own but we adopted two girls."

The next two characteristics explore narrators' attitudes toward hard work and religion. A narrator mentioning that "when the Lord gives us an extra blessing we divide that also with the church" likely considers religion significant. Similarly, a rancher boasting "I've sold more cattle than any other cattleman in boots" signals that hard work shaped their success. The final two characteristics reflect personality traits, which researchers increasingly extract from large text datasets (see e.g. Pennebaker, Mehl and Niederhoffer, 2003; Mairesse et al., 2007). We focus on agreeableness and empathy. Those with agreeable personalities exhibit friendliness, kindness, cooperativeness, or politeness –as illustrated by the cowboy describing workplace camaraderie: "It's true we razzed on each other a lot, but I never saw a cowboy that wouldn't stand up for another when he needed help." Empathy often manifests through kind actions, exemplified by the narrator and his wife who adopted two young girls.

3.2 Human vs Computer Readings

To compare computer and human readings of the life narratives, we randomly selected 300 narratives for analysis by six human readers and two types of computer 'readers.' The human readers were native English speakers pursuing undergraduate or master's degrees in social sciences. GPT-4, our main large language model, has evolved from obscurity to become a household name in recent years.

Large language models like GPT-4 offer significant advantages in document analysis. Trained on billions of internet documents, they can complete text queries and perform annotation tasks at or above human levels (see e.g. Gilardi, Alizadeh and Kubli, 2023). However, they also have notable limitations, producing illogical, inaccurate, and insensitive output at times (e.g. Ashwin, Chhabra and Rao, 2023; Dell, Forthcoming). Additionally, it is unclear they can parse the antiquated language of older adults interviewed in the 1930s.

Our method for using GPT-4 to record these selected characteristics from the randomly selected 300 life narratives is as follows. First, we break each narrative down into chunks of e.g., 150 words, and feed them into GPT-4 one at a time using Open AI's API. We then submit a query like the following for each chunk.

GPT-4 Instructions

You will be provided with text delimited by triple backticks. Your task is to decide whether this person indicates that hard work is important to them. Give your answer as a yes or a no. If yes, give evidence from the text that indicates that hard work is important to them in a JSON object with "answer" and "evidence" as keys. If no, format your response as a JSON object with "answer" and "evidence" as keys and set evidence as "unknown."

For each chunk, GPT-4 provides a yes-no answer with explanatory text, where 'yes' indicates evidence of the characteristic in question and 'no' indicates its absence. For instance, if a text chunk makes no reference to work, it receives a 'no' for that chunk. We aggregate from the chunk level to the document level using a threshold rule: a document receives a 'yes' if the number of positive chunks exceeds a specified threshold. We calibrate these thresholds question by question across all 300 documents so that the probability the LLM and human readers answer 'yes' is as similar as possible. Note that this approach does not target which documents receive 'yes' scores, only the overall fraction.[9]

[9]Using 150 chunks, for example, this results in thresholds of 1 for parental loss, parental loss when young, and empathy; 2 for poverty ending schooling and hard work; 3 for religion; and 6 for agreeableness.

As a frame of reference, we employ pre-specified dictionaries to extract five of the seven characteristics. Dictionary-based text analysis is common in economics and offers several advantages over LLMs. The Linguistic and Word Count software (LIWC), for example, codes text sentiment using standard word association sets. This method excels in transparency, ease of implementation, and accuracy for concepts defined by unambiguous terms. To measure hard work, for example, we used the LIWC dictionary associated with 'work, drives, and achievements.' We counted the number of words in each document that matched the LIWC dictionary words and computed a score. For 'parent died' and 'parent died when young,' we looked for synonyms of 'mother' or 'father' and 'died' in close proximity (see Appendix for our exact methods). For each document and question, we again used threshold system in the score, and we calibrated the thresholds to make the share of 'yes' scores as similar to the human shares as possible

3.3 Turing Tests of Computer Readings

We examine how the scores from the LLM and dictionary approach differ from those of our human readers. To do so, we develop a variant of the Turing (1950) test that asks whether we can distinguish the LLM's answers from the human answers along certain dimensions. The core idea is to compute the fraction of documents where GPT-4 and the dictionary approach match the modal human score. Since we don't have a 'ground truth' for any of our questions, we use the modal human score as a yardstick. If the LLM finds the modal human score with a frequency within the range of the human readers, we say that it has passed the test. This approach captures the spirit of Turing's original test, which asked whether one could identify which answers came from a human and which came from a computer. Passing our test for a particular question means that GPT's answers do not stand out relative to the other human readers for that question. In other words, for that question, GPT can be treated as just another human reader of these documents.

To help illustrate how our method works, Table 4 provides some examples scores from a selected set of narratives for the question of whether hard work is important to the narrator. The first six columns represent the human readers' scores. These are followed by the modal human score, GPT's score, and the dictionary-based score. Both GPT-4 and the dictionary approach agree with the modal human score in Minnie Marshall's case. For Elmo Acosta, the modal human score is again 1, as is GPT-4's score. However, the dictionary score is 0, as Elmo's descriptions of work (focused on his contributions in city government) didn't intersect enough the LIWC work words. There is agreement for Jeff Waggoner, who received unanimous 1 scores, and for Melinda Parker (a narrative about a woman who is sick in bed), whom everyone scored 0.

We implement the Turing Test with all scored documents as follows. For each question and nar-

Table 4: Sample Scores: Hard Work is Important

Life History	Human Reader Scores (1)	(2)	(3)	(4)	(5)	(6)	Modal Human Score	GPT Score	Dictionary Score
Minnie Marshall	0	1	1	0	1	1	1	1	1
Elmo Acosta	1	1	1	1	0	1	1	1	0
Jeff Waggoner	1	1	0	1	0	1	1	1	1
Melinda Parker	0	0	0	0	0	0	0	0	0

Note: This table reports a sample of scores for the question of whether the narrator indicates that hard work is important, where 1 means yes and 0 means no. The first six data columns represent the scores of the human readers. The last three columns are the modal human score, the score of GPT-4, and the dictionary score.

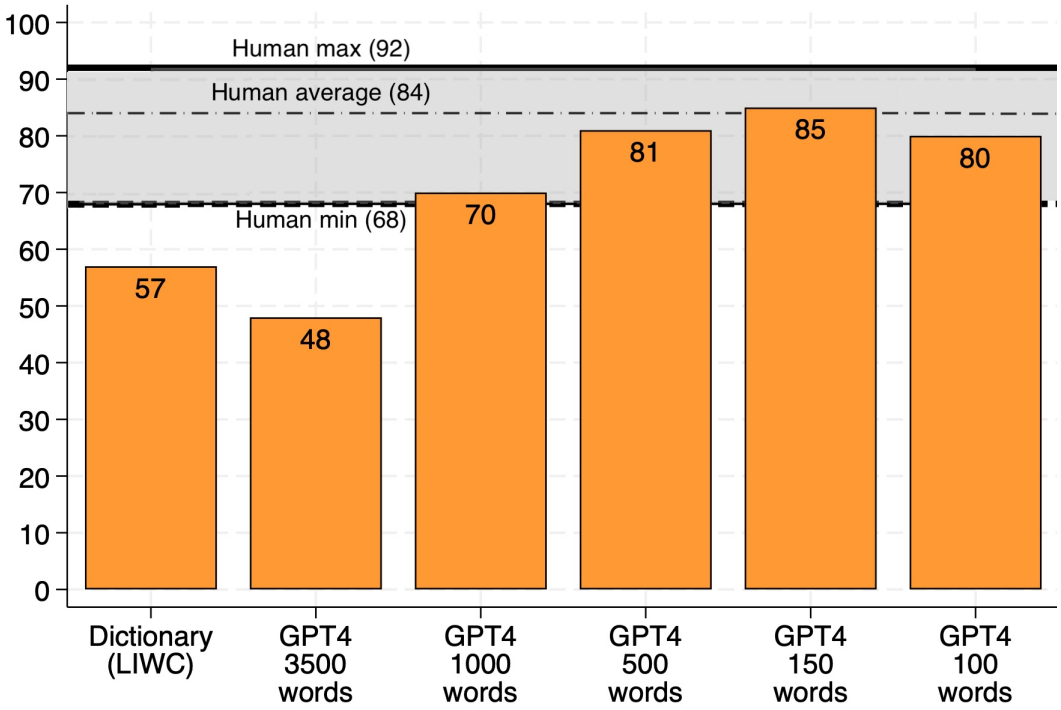
rative, we first determine whether each human's score matches the modal score, as shown in the examples above. We then ask whether the scores of GPT-4 and the dictionary approach are the same as the modal human score. Repeating this for all narratives, we compute the accuracy of each human relative to the modal human score. We then calculate the same accuracy measures for GPT and the dictionary approach. Our key question is whether the accuracy of GPT-4 and the dictionary method falls within the range of human reader accuracies.

Figure 3 reports the results of the Turing tests for whether hard work was important for the narrator. The horizontal lines indicate the range of human reader accuracy: from a minimum of 68 percent to a maximum of 92 percent, and an average of 84 percent. The dictionary approach fails the Turing test, with an accuracy of 57 percent, falling well outside the human range. When GPT-4 is queried with chunks of 3,500 words (or less for shorter narratives), its accuracy is 48 percent, similar to random guessing. Accuracy rises to 70 percent with a chunk size of 1,000 words, just barely passing the Turing test. GPT-4 reaches its peak accuracy of 83 percent with 150-word chunks, nearly matching average human accuracy. Surprisingly, accuracy drops when using 100-word chunks, revealing that shorter text isn't always better. This exercise demonstrates that LLMs can produce human-like answers to yes-no questions when given appropriately sized inputs.[10]

Figure 4 reports the results of the Turing Tests for yes-no questions across our higher-level topics. For each question, we measure GPT-4's accuracy using 150-word chunks. The comparison includes 16,200 human answers (9 questions x 6 readers x 300 narratives) and 2,700 LLM answers (9 questions x 300 narratives). The left-most (blue) bars display the average human accuracy and

[10]We have found similar results for Claude AI (Opus), released March 2024, which also passes the Turing Test when given 150 word chunks but fails when given chunks of 3,500 words. An alternative (and less demanding) version of the Turing Test compares the LLM's scores to the mode including the LLM's own score. Our approach passes that test for all of the characteristics; see Figure A8.

Figure 3: Turing Tests for "Hard Work is Important"



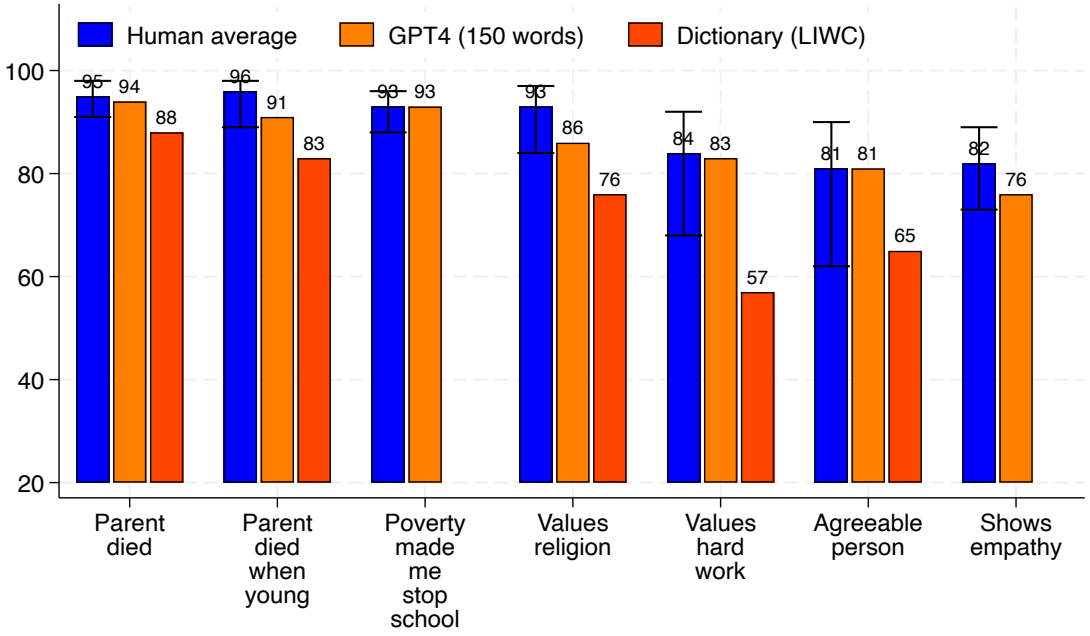
Note: This figure compares the accuracy of the "hard work is important" scores for the 300 randomly selected narratives by 6 human readers, a dictionary method, and GPT under various specifications. Accuracy is relative to the modal human score. The dashed lines represent the minimum, maximum, and average accuracy across the human readers.

accuracy range for each question. For example, with 'parent died,' the average human reader matched the modal human score 95 percent of the time, with accuracy ranging from 91 to 98 percent. The middle (orange) bars show GPT-4's accuracy, while the right-most (red) bars indicate the dictionary method's accuracy.

Looking across Figure 4, one can see that GPT-4 passes the Turing test for every question, finding the modal human score at rates comparable to human readers. The dictionary approach performs reasonably in some cases but fails the Turing test for all characteristics except agreeableness, where it barely enters the range of human scores. For more complex characteristics, human readers show greater disagreement, as evidenced by their lower average accuracy when moving from 'parent died' to 'showing empathy.' GPT-4's accuracy follows the same pattern, declining for these complex characteristics while remaining within the range of human scores. This comparison suggests that GPT-4 can read these life narratives with accuracy similar to our human research assistants.[11]

[11] One potential concern is that the human readers got their answers from an LLM. We cross-checked the human scores for "hard work is important" for the first 100 documents with the website GPTzero, which reports the probability

Figure 4: Turing Tests: Computer vs Human Readings



Note: This figures compares the accuracy of six human readers, GPT-4, and a dictionary method relative to the modal human score for each of the higher-level characteristics. The left-most bars represent the human average; the middle bars represents GPT-4, and the right-most bars are for the dictionary method.

3.4 Results – All Life Histories

Given the successful Turing Test results, we used GPT-4 to answer the same questions for all 1,335 life histories. At this scale, the LLM offers significant savings in both time and money compared to human readers. Assuming a human reader needs one hour per document to answer these questions, scoring the full set would take approximately 9 months. The same task, using GPT-4 with 150-word chunks, took only days and cost approximately $207.[12]

We report the frequency of each characteristic in the life histories in panel (a) of Figure 5. Around one third of narratives mention the death of a parent, and the bulk of these are when the narrator was young. Two thirds indicate that hard work is important; only a minority signal the same about religion. A strong majority show signs of having an agreeable personality, consistent with the findings of the psychology literature, and having showed empathy for others.

that a given piece of text was written by a LLM. None of the human readers had answers with probabilities above 5 percent. Figure A7 reports GPTzero's probability that the human and GPT answers to questions about life satisfaction were written by an LLM. These are always above 80 percent for GPT-4 and below 7 percent for the human readers.

[12]It took a total of 6,897,945 tokens, which, as of summer 2023, cost $0.03 per thousand tokens.

Panel (b) of Figure 5 summarizes differences in the frequency of these characteristics across population sub-groups using odds ratios. In this panel, and similar ones that follow, we control for log word count and interviewer fixed effects. These controls address potential concerns that observed differences might stem from variations in narrative length or interviewer style. However, it is important to note that interviewer fixed effects, while useful controls, substantially limit our observed variation across narrators since they effectively act as detailed location-of-interview controls. The writers would not have had the budget to conduct interviews over a wide geographic range, and most ended up writing about people close to home. For instance, interviewer Mari Tomasi conducted all 19 of her interviews in her hometown of Barre, Vermont.

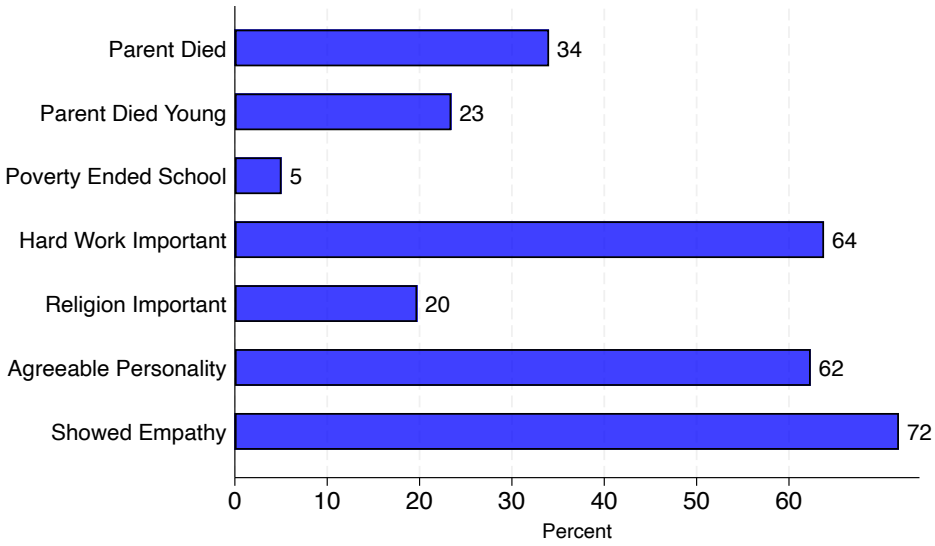
Looking down the first column in panel (b), one can see that a narrator's gender plays an important role in determining the focus of their life history. Parental deaths are unlikely to be very different across males and females in practice, but discussion of parental death is significantly more likely among female narrators. The value of 1.65 in the first row, first column, means that women are 1.65 times as likely as males to describe the death of a parent in their narratives (with the three stars indicating a statistically significant difference at the 1-percent level). This finding will resurface in the analysis to follow. Women were more than four times as likely to say that poverty kept them out of school and more than twice as likely to highlight the importance of religion. Consistent with findings from the psychology literature, women are more likely to have an agreeable personality (see e.g. Schmitt et al., 2008).[13]

The characteristics' incidence also depends on race, occupation, urban-rural status, and whether individuals were immigrants or internal migrants. Religion being important is about fifty percent more likely for non-whites, consistent with their greater religiosity in this period. Those in high-SES occupations – defined as a Duncan SEI score above 20 – were 60 percent less likely to indicate that poverty made them quit school. Many high-SES jobs in this era would have required literacy and numeracy skills that one could not acquire in just a few years of grade school. But those lower on the occupational ladder were around twice as likely to signal the importance of hard work, consistent with the higher average work hours of lower-wage workers in the United States in this period (Costa, 2000), and in lower-income economies more generally (Bick, Fuchs-Schündeln and Lagakos, 2018). Immigrants were almost four times as likely to signal the importance of hard work, consistent with their likely motives to immigrate in the first place. Internal migrants were half as likely to say that poverty ended their schooling, signaling that their destinations may have offered better educational prospects.

[13] We experimented with automated topic identification

Figure 5: Presence of Representative Characteristics

(a) Frequency



(b) Odds Ratios of Characteristics by Group

Source	Female	Non-White	High-SES	Immigrant	Urban	Mover
Parent Died	1.65***	1.09	1.01	0.90	1.19	1.18
Parent Died Young	1.45*	0.91	0.77	0.72	1.33	1.29
Poverty Ended School	4.28***	1.70	0.40**	0.42	1.06	0.48**
Hard Work Important	0.70*	1.45	0.55***	3.84***	0.85	1.27
Religion Important	2.39***	1.56	1.65*	0.89	0.93	1.08
Agreeable Personality	1.73***	1.66	1.48**	0.97	1.06	1.02
Showed Empathy	1.87***	0.64	1.20	0.70	1.11	0.85

< 0.5 | $0.5 - 0.8$ | $0.8 - 1.2$ | $1.2 - 2$ | > 2

Note: Panel 5a reports the frequency of each characteristic in the life histories. Each cell in Panel 5b reports the odds ratio for each characteristic and subgroup, defined as the probability that narrators within each group mention a characteristic relative to the probability those *outside* the group mention it. The groups are women (relative to men), non-whites (relative to whites), high-SES (Duncan's SEI above 20, relative to 20 or below), immigrants (relative to native-born), urban (relative to rural), and movers (relative to stayers). The odds ratios, derived from logistic regressions, control for interviewer fixed effects and the log number of words. Figure A18 reports the corresponding odds ratios not controlling for interviewer fixed effects. Stars represent the range of p-values of the tests that the likelihoods are the same based on robust standard errors, where *** means 1%, ** means 5%, and * means 10%.

4 Sources of Happiness and Meaning

> ”the only thing that really matters in life are your relationships to other people.”
>
> - George Vaillant, *Harvard Study of Adult Development*

Life is more than a sequence of events, both meaningful and trivial. In describing their lives, people become “interpreters of their own life histories” and tell ”stories they had chosen over time as giving meaning to their lives” (Banks, 1980, p. xii). A large literature in psychology has discussed “meaningfulness” (Steger, 2012; King and Hicks, 2021; Heintzelman and King, 2014). It emphasizes three main aspects – purpose, the ultimate motive or aim of one’s life, as in Frankl (1963); coherence or comprehension, the extent to which multiple activities are understood to align with this aim, (Antonovsky, 1993), and significance (making a difference relative to the stated aim). Drawing on the rich set of narratives in the American Life Corpus, we use a combination of human and LLM readings to extract the sources of happiness and meaning that the narrators report. We then summarize these sources across all narrators and examine the rich heterogeneity across demographic and economic sub-groups.

4.1 Open-Ended Questions to Humans and LLMs

To understand the sources of meaning in the lives of the narrators, we begin by analyzing 140 randomly selected life histories using both human readers and LLMs. Four independent human readers examined each narrative, identifying topics that brought the narrator happiness, satisfaction, meaning, or purpose. The readers categorized these themes using a list of 25 common topics (created by one of the readers) plus a catch-all ‘other’ category. We elicited LLM responses using one or several calls to LLMs, and techniques from the recent literature on LLMs, as we explain below. Our simplest LLM query was as follows:

GPT4 Instructions: Sources of Happiness, Satisfaction, Meaning, or Purpose

You will be provided with text delimited by triple backticks. Please identify topics that bring happiness, satisfaction, meaning, or purpose into the life of the narrator. For each topic, please answer in a JSON object with **'topic_name'**, **'reasoning'**, and **'evidence'** as keys.

'topic_name' should be populated with a detailed description of the topic.

'reasoning' should include a short paragraph with your reasoning for the topic.

'evidence' should include a quote from the text that supports your reasoning.

In answering this question, GPT-4 identified hundreds of different topics across the 140 narratives. Some were familiar, such "providing for children's well-being," or "influence of mother," which emphasize the bonds of family. However, many topics did not fit neatly into a standard set of boxes. Some respondents cherished the serenity of living close to nature or caring for natural resources such as forests or livestock. Others were proudest of the adventures they lived through when young, such as run-ins with Billy the Kid, pioneering trips west, or surviving shipwrecks or fires. Faith in God was the guiding force behind some lives, and many found meaning in the educational opportunities they or their offspring received. Some found purpose in their resilience in the face of seemingly insurmountable challenges to their health or economic prospects. The example below presents one specific response from an LLM, taken from narrative of one Elizabeth Vanvderville Darby, interviewed at her home in South Carolina at the age of 84 years.

Example GPT-4 Output: Sources of Happiness, Satisfaction, Meaning, or Purpose

topic_name: "Education at the Moravian School",

justification: "Elizabeth Vanderville Darby found great satisfaction and meaning in her education at the Moravian School. The positive environment, characterized by kind and conscientious teachers, provided her with knowledge and wisdom that she valued throughout her life. this experience was a significant and uplifting part of her life, contributing to her personal growth and development."

evidence: "In my ninth year I was sent to the moravian school, and it's a god's blessing I was. The school was in charge of a Mr. De Schweinitz. The teachers were kind in disposition, conscientious and thorough in their training, and the knowledge and wisdom I acquired there have been useful all my life."

Answers from LLMs typically sound persuasive, and this example is no exception. How trustworthy they are is a different question. We approach the accuracy of the LLM answers in the same way as in the previous section, using a version of Turing's Indistinguishability Test. As we will demon-

strate below, these open-ended questions about sources of meaning pose even greater challenges for LLMs than yes-no queries.

Many of the topics identified were quite similar to one another. For example, "providing for children's well-being" closely resembles "family and providing for children," and "community support and generosity" overlaps with "neighborly kindness and community support." To facilitate our analysis and presentation, we group these LLM-identified topics into eight broader categories (plus an 'Other' category). Table 5 presents these categories along with representative evidence from the narratives.

4.2 Techniques from Computer Science and the Turing Test

To elicit accurate answers from the LLMs, we borrow from the computer science literature on agentic workflow. This involves multiple calls to LLMs and can lead to higher quality output than any single call could provide on its own (Singh et al., 2024). In our case, we consider up to three calls to LLMs for each query.[14] The first asks one LLM a simpler yes-no question: whether there is *anything* that brought happiness, satisfaction, meaning, or purpose into the life of the narrator. The second asks the open-ended question above about the sources of life meaning, but only if the first recorded a positive answer. If not, we record no sources other OpenAI and Anthropic models.[15]

The example below illustrates our multistage assessment workflow through the life history of Mr. Bradley Kennelly, interviewed in 1939 in Jacksonville, Florida. The item and quote identified by the first run of GPT-4 fails the quality check by the second LLM call, so we discard "Career in the Naval Store Business" as a source of life meaning for this narrator.

[14] In contrast to full agentic workflows, there are no loops with agents autonomously making decisions on when to pass back new "work orders" to another LLM, which is why we refer to our technique as "multi-stage assessment".

[15] We have also experimented with a fourth call to an LLM solely to check that the quote used as "evidence" actually appears in the document in question. In all cases we found that the quote does appear.

Table 5: Sources of Happiness, Satisfaction, Meaning and Purpose

Category	Example
Adventure / Exploration / Excitement	"I saw several men riding horseback, coming towards the house. I did not have time to warn Billy [the Kid] that someone was coming but he and mother saw them."
Community / Service	"There is nothing I enjoy better than the stockmen's convention at Fort Worth. Last year I saw two hundred or more of the old boys on given name terms. 'Hey! John," or 'Hello Bill,' could be heard on every side."
Education / Learning	"Five of my oldest children are in New York. All have good jobs and a high school education."
Family / Relationships	"Let me tell you, I sho did have a good husband. He made $15.00 a week wuking at de Holman Building, and evvy Sadday night he fetched evvy last penny of dat money straight home and laid it in my lap.'
Nature / Outdoor Activities	"Mr. Stevenson takes a great delight in hunting, and tells of many thrilling hunting trips."
Religious / Spiritual Beliefs	"Since I haven't been able to walk for over a year I don't go to church, but I listen to good sermons over my radio."
Resilience / Overcoming Challenges	"She never grieved over her blindness but went right along cheerful and working all the time."
Work / Career	"Mr. Botsford produced an ancient wooden instrument of which he was obviously proud. 'I could take this clock today, and put a pendulum and a dial on it and start it and I bet it'd keep perfect time.'"

Example GPT-4 Output and Quality Check

topic_name: "Career in the Naval Store Business",
justification: "Mr. Kennelly expresses his dedication and passion for his job in the naval store business, despite the challenges he faces. This career brings him a sense of purpose and fulfillment."
evidence: "I have been connected in this present capacity for nine years and I can honestly say that I've never seen business so slow in this line ever before."

---

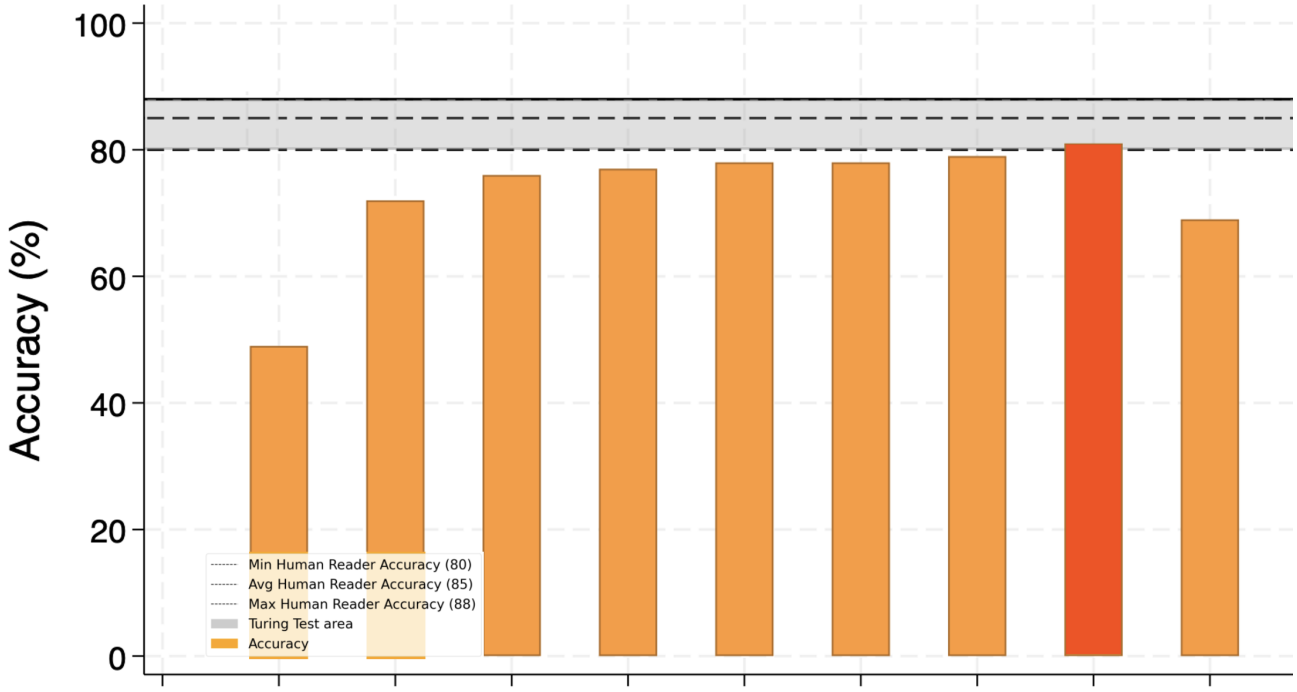
quality_check: "No."
feedback: "The evidence does not clearly state that Mr. Kennelly's career in the naval store business brings him happiness, purpose, or meaning. His statement about the business being slow does not necessarily indicate dedication or passion."

As a way of summarizing the added value of each technique we employ, Figure 6 reports each technique's accuracy, relative to the modal human score, in identifying work/career as a source of happiness and meaning. Each bar represents the result of making one or more specific calls to an LLM for each story, for the human audit sample of 140. As before we also compute the accuracy of each human relative to the modal human score. The minimum and maximum human accuracies (80 and 88 percent) are plotted as the solid and dashed lines at the top of the figure. The average accuracy (85 percent) is the horizontal dashed line in the middle.

The first bar on the left represents the simplest and most obvious way to query an LLM with an open-ended question: by just asking it and feeding in the entire life history. Unfortunately, this approach fails miserably. Accuracy is just 48 percent, which is well outside the human range, and no better than chance. The second bar shows what happens when we break up the text into 500 word chunks, as in the previous section. We include any topic found in any one of the chunks as a source of meaning for that narrator. This leads to a sizable increase in accuracy up to 69 percent. As with the yes-no questions of Section 3, chunking is a simple way to significantly improve LLM output. Unlike the yes-no questions, however, chunking alone is not enough to break into the human range and pass the Turing Test.

The third bar shows the effect of moving from GPT-4 to GPT-4o, a more recent iteration of OpenAI's LLM models (released in May 2024). The fourth and fifth bars reflect the effect of more detailed guidance to the LLM and an enhanced prompt with examples (see Appendix A for the exact prompts). Each makes incremental progress but the LLM continues to fail the Turing Test. The sixth bar introduces the second LLM call that provides a quality check on the first, as in the example of Mr. Kennelly above. This leads to a marginal improvement in accuracy but still results

Figure 6: Turing Tests for Work/Career as a Source of Meaning



Prompt type	Simple	Simple	Simple	Simple	**Enhanced**	Enhanced	Enhanced	Enhanced	Enhanced
Guidelines	No	No	No	**Yes**	Yes	Yes	Yes	Yes	Yes
Model	GPT4	GPT4	**GPT4o**	GPT4o	GPT4o	GPT4o	GPT4o	GPT4o	GPT4o
Chunking	No	**500**	500	500	500	500	500	500	500
Quality Check	No	No	No	No	No	**Yes**	Yes	Yes	Yes
Confidence Scores	No	No	No	No	No	No	**>=7**	**>=8**	**>=9**

Note: This figure reports the accuracy of the LLM output, relative to the four human readers in identifying work/career as a source of life happiness and meaning. Each bar represents one specification.

in failure.

The red bar on Figure 6 represents the first passing of the Turing test for the open-ended question. In addition to all of the above, in the specification represented by this bar (and the two adjacent ones) we ask the first LLM to report its confidence in its own answer on a scale of 1 to 10. We then create a cutoff for each value on the scale, ignoring any answers below the cutoff, and compute recall and precision for each cutoff. The red bar is the value that results in precision and recall being as similar as possible. This turns out to occur at a cutoff of 8. The other bars are for cutoffs 7 and 9, and both result in inferior accuracy (and failure). The cutoff of 9 fails because the LLM output is assessed too cautiously – we throw out many topics that the humans actually identified, resulting in low recall.

Table 6: Turing Tests: Computer vs Human Readings

Percent of Narratives with Modal Human Score		
	Human Average [min, max]	GPT 4o (500 words)
Any Happiness, Satisfaction, Meaning, or Purpose	76 [65, 83]	81
Family	85 [78, 91]	77
Work and Career	85 [80, 88]	81
Religious and Spiritual Beliefs	94 [92, 95]	94
Education and Learning	94 [88, 97]	89

Note: This table compares the readings of the 130 randomly selected narratives by 3 human readers and GPT-4o. For Happiness, Satisfaction, Meaning or Purpose, the score is either 0 (e.g. unhappy or not sure) or 1 (this is a happy narrative). The remaining characteristics are from open ended questions to GPT about what brought happiness, satisfaction, meaning, or purpose to the narrator, categorized into groups comparable with categories selected by humans. The first column reports the average percentage of humans agreeing with the modal score, and below it, the minimum and maximum percent across the RAs. The second column reports the share of the time that GPT find the modal score.

The upshot of these results is that LLMs can pass a Turing Test for sophisticated open-ended questions, but only with sufficient guidance from human users. For our documents, we need a mix of chunking, multiple calls to LLMs for quality checks, and judicious use of confidence thresholds using the LLM reports of confidence in its own answers. Table 6 reports the results of the Turing Tests for the initial LLM query about happiness and four sources of happiness and meaning that could be directly compared to the human reader answers: family, work/career, religion, and education. In all cases, we use our preferred specification corresponding to the orange bar in Figure 6. This method results in LLM answers that pass the Turing test for education and religion and land narrowly below the human range for family.[16]

[16]It passes on all categories in the less demanding case when the LLM's answers are counted when computing the modal score; see Appendix Table A.1.

4.3 Sources of Life Meaning

Figure 7, Panel (a), reports the share of narrators who mention each source of meaning. Each document mentions between zero and nine sources, with a mode of two.[17] Work and career emerges as the most common category, representing significant sources of happiness, satisfaction, meaning, or purpose in 49 percent of the documents. Community and service follows at 45 percent, with family and relationships next at 39 percent. The central importance of work among our narrators contrasts with the famous Harvard Study of Adult Development (Vaillant, 2012, 2008), which emphasizes the paramount importance of relationships with other people (see also the recent evidence of Geiecke and Jaravel (2024)). Less prominent sources of meaning include nature and outdoor activities (present in 16 percent of narratives), adventure and exploration (14 percent), resilience and overcoming challenges (13 percent), and education and learning (13 percent). Religion and spirituality, plus the residual 'other' category, appear least frequently at 11 and 10 percent, respectively. The overall patterns remain similar when we weigh the frequencies to mimic the elderly US population. See Figure A9. Moreover, in Figure A13 we explore how the sources of meaning of life differ between movers and stayers further distinguishing among movers to those moving west (or moving to an urban destination).

We find the emphasis on work and career as a source of meaning in life particularly intriguing. For *Homo Economicus*, work is but a means to an end. For example, Ralph Lifshitz, working at the chicken market in New York said "It's not exactly a business I would want to go into if I had the choice," lamenting that he "never thought I'd end up here." Yet for around half of the elderly narrators looking back on their lives, their work was enriching enough to describe it prominently and favorably in their life stories. Many such individuals found a sense of purpose in helping others. The life story of Elmo Acosta, and his commitment to public service, is just one example. Mariah Jackson of Athens, Georgia, found her mission in her vocation as a midwife, devoting herself to the women she helped safely bring children into the world. Others found deep contentment in their freedom and self-sufficiency on the job, such as Jeff Waggoner, who roamed the ranges answering to no one on a daily basis. Still others felt keen satisfaction from their career accomplishments or the valuable skills they acquired. R.W. Wishart, whose narrative is titled 'Turpentine Man,' achieved such stature in the turpentine industry of his native Florida that the government of Haiti brought him in to help them develop their own industry.

Sharp differences emerge when comparing what brings happiness and meaning across sub-groups

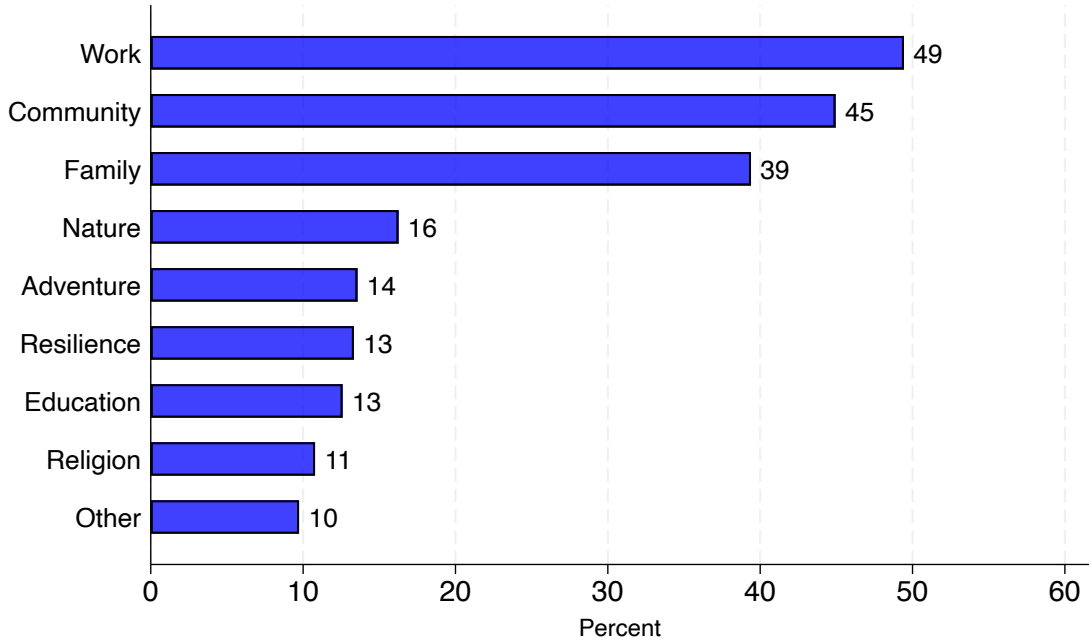
[17] In 34% of our narratives, no source of life satisfaction is identified. This number is at the high end compared to results from the World Value Surveys, but not out of range compared with countries like Japan, Spain, or France in 2007 (Oishi and Diener, 2014). As it will become apparent in Sections 6 and 7, the lives of these individuals who did not find a source of meaning in their lives are replete with negative turning points that create a rather bleak life profile.

of narrators. These appear in Figure 7, panel (b), which presents odds ratios by sub-group (see Appendix Figure A12 for the raw frequencies by sub-group). Family and community were significantly more likely to be mentioned as a source of meaning by women, consistent with widely-held wisdom about gender roles in the late 19th and early 20th century. More surprisingly, resilience and religion were also markedly more common uplifting themes for women than for men. This result relates to the findings in Section 3. There, we found that women were more likely to highlight a parental death or that poverty ended their schooling. It suggests that women suffered more from negative shocks and turned to religion more as a result. One area of agreement was work and career, which women were only slightly less likely to mention as a source of meaning (and with a statistically insignificant difference between men and women).

Those in high-SES jobs found more meaning in their work, suggesting that better jobs are not just higher paying but also offering better non-wage amenities. Immigrants were also significantly more likely to find meaning in work than natives, consistent with their motives for leaving their homelands in the first place. Among non-whites, faith showed up more strongly as a source of meaning, consistent with their greater religiosity in this era. Those who moved within the country were markedly more likely to be uplifted by education but also adventure than those who stayed put.

Figure 7: Sources of Happiness, Satisfaction, Meaning, or Purpose

(a) Frequency



(b) Odds Ratios by Group

Source	Female	Non-White	High-SES	Immigrant	Urban	Mover
Work	0.87	0.64	1.59**	1.96**	0.88	1.43*
Community	1.45**	1.31	1.23	1.64*	1.13	1.21
Family	1.85***	0.52**	1.47**	1.18	0.83	1.29
Nature	1.05	1.19	1.21	1.21	0.81	0.98
Adventure	0.93	0.64	1.46	1.32	1.05	1.68*
Resilience	1.92**	0.27**	1.37	1.43	0.69	1.41
Education	1.28	1.11	2.00**	1.18	1.14	1.89**
Religion	2.09**	1.99*	1.09	0.94	1.14	1.64*

< 0.5 | $0.5 - 0.8$ | $0.8 - 1.2$ | $1.2 - 2$ | > 2

Note: Panel 7a reports the frequency of each source of meaning in the life histories. Each cell in Panel 7b reports the odds ratio for each source and sub-group, defined as the odds that narrators within each group derive meaning from a given source, compared to the odds of those *outside* the group finding meaning in it. The groups are women (relative to men), non-whites (relative to whites), high-SES (Duncan's SEI above 20, relative to 20 or below), immigrants (relative to native-born), urban (relative to rural), and movers (relative to stayers). The odds ratios, coming from logistic regression, control for interviewer fixed effects and the log number of words. Figure A19 reports the corresponding odds ratios not controlling for interviewer fixed effects. Stars represent the range of p-values of the tests that the likelihoods are the same based on robust standard errors, where *** means 1%, ** means 5%, and * means 10%.

4.4 Ancestral Roots of Life Meaning

While a comprehensive analysis of what drives the variation in life meaning across narrators lies beyond this study's reach, we share findings about one promising predictor: the narrator's ancestral background. Our inquiry is motivated by the observation that ancestral stories, which differ across societies, may shape how one interacts and perceives the world around them. Among our narrators, we can trace the country of origin for 384 out of 1335 individuals. Among those, 75 percent were born in the United States and the rest are first-generation immigrants. We have a total of 26 ancestries spanning mostly European origins. The largest group is German, followed by Irish, English, and Scottish, in that order.

Table 7: Ancestry and Sources of Meaning

Dependent Variable	Family	Work	Community and Service	Adventure	Nature	Religion	Resilience	Learning
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A								
Challenge Overcome in Folklore	-0.00	-0.00	-0.01	-0.00	-0.03**	-0.01	0.04***	0.01
	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)
Adj. R^2	0.160	0.101	0.046	0.121	0.027	0.071	0.049	0.064
N	384	384	384	384	384	384	384	384
Dep. Var. Mean	0.56	0.71	0.70	0.22	0.22	0.12	0.19	0.16
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel B								
Male Bias in Folklore	0.01	0.03*	-0.02	-0.00	-0.00	-0.05**	-0.00	-0.03
	(0.04)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)
Female	0.17***	-0.15***	0.04	-0.12***	0.05	0.05	0.08***	0.09**
	(0.06)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.02)	(0.04)
Male Bias in Folklore × Female	-0.01	-0.05**	0.04	-0.01	-0.06	0.03	-0.06**	-0.02
	(0.04)	(0.02)	(0.04)	(0.02)	(0.04)	(0.03)	(0.02)	(0.05)
Adj. R^2	0.158	0.101	0.044	0.119	0.025	0.078	0.036	0.069
N	384	384	384	384	384	384	384	384
Dep. Var. Mean	0.56	0.71	0.70	0.22	0.22	0.12	0.19	0.16
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: In both Panels we control for the narrator's gender, age-bin fixed effects, and log word count. The sample is restricted to narrators who have found meaning in life. Those with American ancestry (irrespective of race) are excluded. The variables "Challenge Overcome in Folklore" and "Male Bias in Folklore" are standardized. Standard errors are clustered at the ancestry level.

To capture ancestral narratives, we leverage information from the folklore country catalogue constructed by Michalopoulos and Xue (2021). We focus on two themes. The first reflects how often

the protagonists in the country's folkloric motifs overcome challenges and manage to get out of difficult situations. The second dimension reflects the degree of stereotypical gender norms as manifested in the share of motifs where men are depicted as more dominant, less submissive, less engaged in domestic affairs, and more physically active than women in the country's oral tradition. We call this male bias in folklore. To abstract from regional differences in the meaning of life, we add region fixed effects in all specifications.

In Panel A of Table 7 we explore how ancestral stories featuring successful resolutions of challenges translate into the various sources of life satisfaction. Among the 9 different categories people have found meaning in, one stands out in relation to challenge-resolving folklore. People whose ancestral stories feature more of the latter are systematically more likely to have found meaning in life by overcoming challenges. A one-standard-deviation increase in such motifs, i.e., tracing one's roots to Sweden instead of Italy, increases the likelihood that the narrator will describe how much (s)he cherished successfully navigating a difficult time in their lives by 4 percentage points (relative to a baseline probability of 19 percent). Besides resilience, having found meaning in nature is less common among people whose folklore has a relative abundance of challenging but successful encounters.

In Panel B, we explore the role of male bias in differentially shaping the sources of life satisfaction across genders. To do so, we interact the country of origin male bias in the oral tradition with a female indicator. Women are 15 percent less likely to derive meaning from work compared to men. Furthermore, females tracing their ancestry in oral traditions that depict them as less independent and more engaged in domestic affairs relative to males are an additional 5 percent less likely to highlight that their career brought fulfillment (a one-standard-deviation increase in male bias is equivalent to comparing Irish to Portuguese folklore). Females from male-biased traditions are also less likely to have found meaning in overcoming challenges in their lives. In contrast, the same increase in male bias increases the likelihood that males found meaning in work by 3 percent. These patterns suggest that ancestral stories may provide a point of reference on where meaning in life is (not) to be found, even for people who have long left their ancestral homes living in the same region.

5 Critical Junctures in Life

Which turning points stand out to people look back at their lives? The trajectory of every human life is defined by a combination of decisions, conscious or unconscious, and external shocks. Henry Kissinger, in his BA thesis, put it eloquently:

> In the life of every person there comes a point when he realizes that out of all the seemingly limitless possibilities of his youth he has in fact become one actuality. No longer is life a broad plain with forests and mountains beckoning all-around, but it becomes apparent that one's journey across the meadows has indeed followed a regular path... (Kissinger, 1950)

Nearly every individual in the American Life Histories made reference to memorable decisions or events that shaped their story. Many were familial milestones, looked back upon fondly, such as when narrator Cora Brauscom recalled meeting her husband: "For a woman my age, I got awful flustered ... We were married so quick I hardly knew how it all was, it fairly took my breath away; my head just swam." Others were perilous episodes of great uncertainty, like when W.E. Oglseby moved west in a wagon caravan along with 17 other families: "All the people had sold more or less property and had a little money, not much, because in those days real estate did not sell for much. They all loaded their personal effects into covered wagons and started for Texas, with high hopes to do better."

Using similar methodology as we described in the previous section, we recorded all of these critical junctures for each narrative using the following prompt.

GPT-4 Instructions: Critical Junctures

Identify and describe the critical junctures in this person's life. List:

(a) The critical juncture

(b) Evidence, citing the supporting text using one or more sentences

(c) Explanation, explain your reasoning, one step at a time.

For each critical juncture, please answer in a JSON object with **'critical_juncture'**, **'evidence'**, and **'explanation'** as keys.

As one instructive example, GPT-4 listed three critical junctures in the life of Alcide Savoie, a granite cutter. One occurred when he moved to Barre, Vermont, from Canada. The second was when he began work in the granite mills. The third was the following:

Example GPT-4 Output: Critical Juncture

"**critical_juncture**": "Moving into Mrs. LaCrosse's boarding house",
"**evidence**": "That's how I met my landlady, at a supper the St. Anne's Society was giving. The society was just for married French women. A Catholic Society. Mrs. LaCrosse did the cooking that night and we all thought it was fine. Her husband had died of stonecutter's T. B. the month before, she told us she was going to invest her insurance money in a good, plain house and take in boarders and roomers. Four of us moved in the next week.",
"**explanation**": "Moving into Mrs. LaCrosse's boarding house was a critical juncture in Savoie's life because it provided him with a sense of community and a place to call home for the next ten years. It also marked a shift in his social life, as he began to form close relationships with his fellow boarders."

The event itself – moving into a boarding house – could easily be called mundane and meaningless. However, for the narrator, it is highly significant not because of the move itself, but because of the changes it engenders in his life. In the boarding house, Savoie connects with the community of boarders and becomes more socially integrated than before. This example demonstrates that, with the right prompting engineering, frontier LLMs like GPT-4 can identify and extract the more subtle critical junctures that alter the path of the narrators' lives.

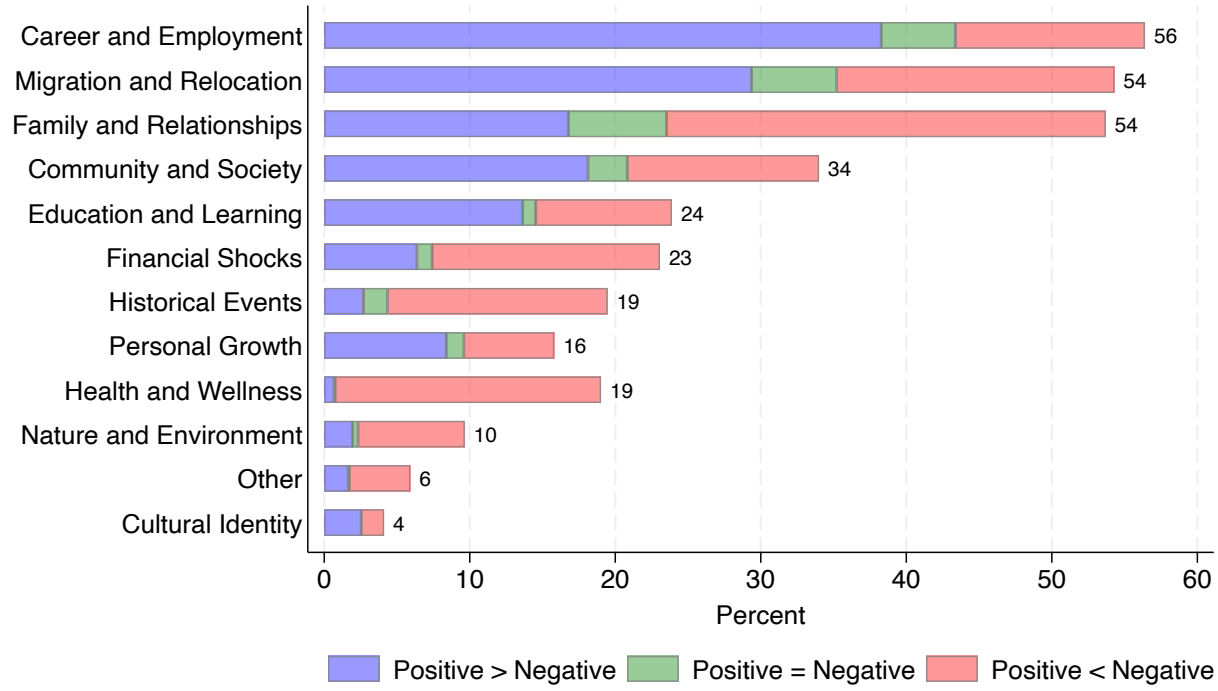
In total we extracted 5,672 critical junctures from our corpus of 1,335 life histories — an average of 4.2 per life narrative. Almost all the life narratives mention 3-4 key turning points. We assign this long list of responses to broader categories, designed to be mutually exclusive. Table 8 lists the broader categories with example text for each. Figure 8 shows the share of life narratives mentioning any event, by category, in our corpus. Career and employment is the single most frequent category. It occurs in 56 percent of life narratives. This is followed by migration and relocation, and by family and relationships, all of which are mentioned in more than half of the documents. Virtually no narrator fails to mention at least one of these categories. The categories education, financial shocks, historical events, and health/wellness are mentioned in about 1 out of every 5 life histories. At the low end, personal growth, nature and the environment, and cultural heritage are mentioned in 5 to 15 percent of cases.

This is not to say that infrequently mentioned categories are not important when they do occur. Weather shocks, for example, could have severe consequences: Jack W. Patterson, a 40-year-old farmer from Uvalde, Texas remembers losing half his flock of sheep when "a terrible drought hit this country and burned things up." Richard Murphy of Fort Worth describes how after the drought of 1918 "during the first part of my (cow) bone gathering, we could gather a wagon load in about 30 or 40 minutes because the ground was simply strewn with carcasses." He then goes on to describe

Table 8: Categories of Critical Junctures

Category	Example Passage
Career / Employment	"Eight years ago I was laid off at the American Tube and Stamping Co. where I worked for a long time."
Community / Society	"The early days of preaching in Oregon have been filled with pleasant memories for my devoted wife and myself. In our evangelistic work we had to travel from house to house...
Cultural Identity / Heritage	"Her search for them led to the discovery of the Native African Union... Through some of her newly found friends, she was able to study and dance with this group..."
Education / Learning	"When still quite young I entered Cedar Grove Academy, near Bamberg, South Carolina."
Family / Relationships	"After my father died, we got along better than we did when he was living. I believe I told you my mother had fine people, well they helped my mother raise us two boys."
Financial Shocks / Economic Changes	"We had plenty of rain in the spring and everything grew fine until along in the middle of the summer when it turned hot and dry up until all we had planted was burned up."
Health / Wellness	"A heavy door closed on the forefinger of my right hand severing it completely."
Historical Events / Context	"I was captured at the battle of Gettysburg in Longstreet's charge and was taken to Fort Delaware, an island of 90 acres of land where the Union prisoners were kept."
Migration / Relocation	"When I was nineteen I told my father I was going to Maine. I wasn't going to work all my life for nothing, and I knew I could get a dollar a day in Maine."
Nature / Environment	"In the earlier days of our childhood we had a terrible time keeping warm. We never knew when a great storm would come up and just how the next day would be."
Personal Growth / Self-Discovery	"I don't know what would have become of me if an old circuit riding doctor, who went from place to place as regular as mail, hadn't of given me an old broke down hoss..."

Figure 8: Frequencies of Positive and Negative Critical Junctures



Note: The figure shows the frequency of each critical juncture across life histories. The colors of each bar indicate the frequency of individuals reporting more positive than negative, more negative than positive or the same number of positive and negative critical junctures.

how they sold 23 carloads of cow bones to Dallas for the Williams Metal Bone Company.

Critical junctures can refer to events where life changed for the better or for worse. Positive or negative junctures may map differentially into how the narrators look back at their lives. To quantify the direction of each turning point we asked GPT-4 to label each critical juncture as either positive or negative. Figure 8 illustrates the share of respondents within each category where the turning points narrated are mainly positive, mainly negative, or balanced. Narrators' descriptions of changes in financial circumstances are more likely to be negative shocks, such as business bankruptcies, harvest failures, etc. Turning points related to family relationships have both positive and negative aspects; negative ones are significantly more common though. Career and employment events are more likely to be uplifting whereas the overwhelming majority of health shocks are negative. Finally, cultural identity and heritage are often discussed in positive terms, and so are community and society, and education and learning. The incidence of a balanced representation of shocks is relatively uncommon suggesting that narrators had relatively homogeneous recollections of how family-related or community-related shocks shaped their lives. The patterns are similar when we weight the frequency to reflect the US elderly population in 1940. See Figure A16.

Figures 9 reports the odds ratios of positive and negative critical junctures across various groups. Important gender differences appear, with women's critical junctures generally painting a more challenging picture compared to men's experiences. Most notably, while women report positive family-related turning points slightly more often than men, they are nearly three times more likely to describe negative family events. These instances are often related to how deaths in the family, like the passing of one's parents, became watershed moments in the lives of female narrators. Women also skew negative on career and migration events.

The disproportionate prevalence of these negative turning points in the lives of women often reflect their roles as kin-keepers. Elizabeth E. Miller is one clear example. "I didn't have no education, but I had the chance of one," she recounts wistfully in her narrative. As the oldest of six children, Elizabeth becomes the primary caregiver for her mother after she takes ill and becomes bed-ridden. A well-off uncle offers to pay for Elizabeth's schooling if she moves in with him closer to the school, but she declines: "I looked at him and said 'Uncle, I can't. I've got to stay with Mother. She needs me.' That was all that was ever said about it." Mrs. T.C. Ingle is another such kin-keeper. Her marriage and family life begin happily enough, but when her husband turns to heavy drinking, she is left to care for their finances and children. "It was up to me to support us," Mrs. Ingle recalls, "because most of what little money he made went for likker."

Critical junctures vary markedly across racial and socioeconomic lines. Non-white individuals were ten times more likely than whites to experience positive turning points tied to cultural identity. They also report positive community-related and personal growth events more than twice as frequently as whites. Interestingly, non-white narratives feature fewer weather-related shocks and significantly fewer migration episodes, both positive and negative. Those in high-SES occupations experience twice the rate of positive career and educational turning points, while facing fewer negative financial events.

The urban-rural divide stands out clearly in the data. Urban narrators report approximately 70 percent more career-related turning points, both positive and negative, likely reflecting the greater job churn of urban labor markets compared with rural employment, particularly family farming. As one might expect, immigrants and migrants also reported systematically different types of turning points. Immigration status correlates with more frequent migration-related turning points, though these show only a slight positive skew. Narrators who migrated during their lifetimes were significantly more likely than non-migrants to report turning points related to career or migration of both a positive and negative nature, with bigger differences for the positive events. Education-related turning points were much more likely for movers and instances of negative personal growth were much less likely.

How do these two key aspects of life narratives — critical junctures and sources of life meaning

Figure 9: Odds Ratios of Positive and Negative Critical Junctures by Sub Group

(a) Positive

Source	Female	Non-White	High-SES	Immigrant	Urban	Mover
Career and Employment	0.23***	0.78	1.84***	1.49	1.68**	1.68***
Migration and Relocation	0.83	0.43**	1.25	3.00***	1.29	2.31***
Family and Relationships	1.26	0.82	0.97	1.34	1.25	1.32
Community and Society	0.91	2.41***	1.26	0.75	1.09	0.96
Education and Learning	0.71	0.75	2.22***	1.13	1.24	1.67**
Financial Shocks	0.98	0.75	0.76	1.44	0.85	0.88
Historical Events	0.64	0.65	0.80	0.14**	1.13	0.59
Health and Wellness	0.67	1.95	3.45	8.04*	0.38	0.80
Personal Growth	0.71	2.91**	0.84	0.53	0.86	1.06
Nature and Environment	0.51	2.94	1.03	1.00	0.79	0.73
Cultural Identity	1.96	9.49***	0.44*	1.74	0.39	0.36*

(b) Negative

Source	Female	Non-White	High-SES	Immigrant	Urban	Mover
Career and Employment	0.52***	1.44	0.96	0.64	1.73**	1.48*
Migration and Relocation	1.46**	0.49**	0.97	2.40***	0.70*	2.06***
Family and Relationships	2.74***	1.45	0.63**	0.70	1.09	0.98
Community and Society	0.83	1.09	1.16	0.92	0.72	0.58**
Education and Learning	1.15	0.62	1.04	0.41	0.56**	0.84
Financial Shocks	1.17	0.85	0.56**	0.60	1.42	0.87
Historical Events	0.99	1.28	0.78	0.58	0.51**	0.79
Health and Wellness	0.88	0.79	1.23	0.70	0.76	0.90
Personal Growth	0.82	1.19	1.24	1.26	0.85	0.25***
Nature and Environment	1.54	0.13**	0.71	1.08	0.59	0.99
Cultural Identity	1.63	0.57	0.59	0.76	0.68	0.36

< 0.5 | $0.5 - 0.8$ | $0.8 - 1.2$ | $1.2 - 2$ | > 2

Note: Each cell in Figure 9 reports the odds ratio for each critical juncture and sub-group, defined as the probability that narrators within each group mention at least one critical juncture (positive in Panel 9a and negative in Panel 9a, compared to the probability of those outside the group mentioning it. The odds ratios, coming from logistic regression, control for interviewer fixed effects and the log number of words. Figures A20 and A21 present the same Panels omitting interviewer FE. Stars represent the range of p-values of the tests that the likelihoods are the same based on robust standard errors, where *** means 1%, ** means 5%, and * means 10%

— interact? Do they mostly capture the same underlying view, or are they distinct aspects of a multifaceted life history? Figure 10 reports odds ratios for various meaning-of-life dimensions, as a function of critical junctures. For example, where a positive career juncture is reported, the respondent is 2.84 times as likely to find meaning in work. The link between critical junctures

related to work and career being a source of life satisfaction is perhaps unsurprising, but other patterns are less obvious: those with a positive career critical juncture are also only 0.57 times as likely to find meaning in religion. Those who found meaning in religiosity and spirituality are nearly twice as likely to mention adverse health shocks consistent with a large literature on how religion helps with coping and more than six times as likely to mention a positive juncture involving personal growth. In fact, those with positive personal growth junctures are significantly more likely to find meaning in every single category (though religion stands out as the highest).

Those facing negative financial shocks were about half as likely to find happiness in work – and also family and community – consistent with spillover effects of job loss or business failure. We return to these issues in Section 7, when we explore how narratives fit into specific clusters based on what brought them meaning and the critical junctures they experienced.

In Appendix Section A.3, we explore the geography of critical junctures and meaning of life across the broad macro US regions.

6 Retrospective Life Satisfaction

Having identified sources of happiness and the critical junctures that shaped people's lives, we now turn to a complementary question: how satisfied were the narrators with their lives? Since they were not asked about this directly, we must infer it from their stories. This section aims to construct a cardinal measure of retrospective happiness / life satisfaction, assess its reliability, and analyze its properties.

6.1 Scoring and Transitivity

We examine how fulfilled each narrator's life was. To this end, we prompt a current frontier model LLM (GPT-4o) to assess the overall life satisfaction of the narrator on a scale ranging from 1 (for overwhelmingly negative life histories) to 7 (for overwhelmingly positive ones).We provide detailed instructions on what to look for at each level (for the exact prompt see Appendix A.1). This yields a distribution of scores with a range of 1 to 6, and an average of 4.31. In other words, in the eyes of the LLM, most people score relatively high in life satisfaction, but there is substantial variability. Before we can use these scores for further analysis, we need to ensure that we capture important variation in the eyes of the narrators. To this end, we conduct two exercises. In the first, we examine whether transitivity of satisfaction assessments holds; in the second, we compare the satisfaction scores from the LLM to human assessments.

Figure 10: Odds Ratios for Meaning of Life by Critical Juncture Presence

Meaning of Life

Positive Critical Junctures	Work	Adventure	Family	Community	Learning	Resilience	Nature	Religion
Career	2.84***	1.71**	0.97	1.26	1.63**	0.92	0.80	0.57**
Migration	1.53**	1.57**	1.27	1.34*	0.94	0.95	1.15	1.07
Family	1.28	0.69	2.00***	1.11	0.68	0.92	0.80	1.11
Community	1.01	0.97	1.04	2.10***	0.63	1.10	0.64*	2.11**
Education	1.15	0.70	1.42*	1.24	5.93***	0.82	1.07	1.53
Financial Shocks	2.31***	1.24	1.32	1.19	0.58	0.84	0.84	1.07
Historical	0.35**	2.10*	0.36**	1.03	0.46	0.85	0.29**	0.18*
Health	0.77	1.45	0.44	1.17	3.00	0.67	1.22	4.32
Personal Growth	2.05**	1.95**	1.21	2.01***	1.54	1.32	1.77*	6.86***
Nature	2.33	3.13*	1.10	1.43	0.34	0.74	2.13	0.39
Culture	0.59	0.86	2.08	1.82	0.47	1.00	2.71	1.00

Meaning of Life

Negative Critical Junctures	Work	Adventure	Family	Community	Learning	Resilience	Nature	Religion
Career	0.80	0.99	0.81	0.55***	1.36	1.25	0.99	1.18
Migration	0.97	1.30	1.02	0.88	0.83	0.90	1.45*	0.83
Family	0.86	1.20	1.31*	0.62***	1.31	1.45*	0.96	0.85
Community	0.56**	0.98	0.70	0.80	0.39**	0.55*	0.93	0.85
Education	0.92	1.07	1.15	0.79	1.03	0.88	0.82	0.70
Financial Shocks	0.52***	0.60	0.56**	0.58**	0.90	0.85	0.68	0.97
Historical	0.64**	0.83	0.85	0.95	1.06	1.20	1.09	0.38**
Health	0.89	0.62	0.96	0.82	0.74	0.97	0.75	1.87**
Personal Growth	0.82	1.08	1.08	0.99	0.63	0.53	0.85	2.18
Nature	1.03	0.90	0.58*	1.05	0.63	2.20**	0.85	1.40
Culture	0.55	0.28	0.35	0.31	0.94	1.00	0.34	0.34

< 0.5 | $0.5 - 0.8$ | $0.8 - 1.2$ | $1.2 - 2$ | > 2

Note: Each cell in Figure 10 reports the odds ratios of a critical juncture with respect to a meaning of life category. Odds ratios are obtained via logistic regression controlling for log number of words and interviewer fixed effects. The first table uses as outcome variable an indicator taking value 1 if narrators report a positive critical juncture. Conversely, the second table focuses on a indicator taking value 1 if they report a negative one. Figure A22 reports the corresponding odds ratios not controlling for interviewer-fixed effects.

Transitivity is a basic test of the quality of comparisons implied by LLM happiness ratings. If narrative A is rated happier than narrative B, and B happier than C, then A should also be happier than C. We can ask an LLM to directly make this comparison, feeding it pairs of stories, and then examining the overall set of scores for inconsistencies. Appendix Table A.2 shows the results for randomly chosen sets of three life narratives. In no case can transitivity of the LLM comparisons be rejected; in some cases, it is not determined by the set of results. For example, in row 2 of the table, Annie Hight is rated above Ella Cox in the direct text comparison, and Ella Cox above Frank Nicho. Finally, Annie Hight appear more satisfied with her life than Frank Nicho, which is consistent. The satisfaction scores of 5, 5, 3 are in line with this relative ordering. The final column shows whether the scores assigned align with the relative ranking from the direct comparisons. Only in 3 out of 60 pairwise comparisons is there an inconsistency between the relative ranking of the direct, bilateral comparison and the set of two LLM scores derived from our general prompt.

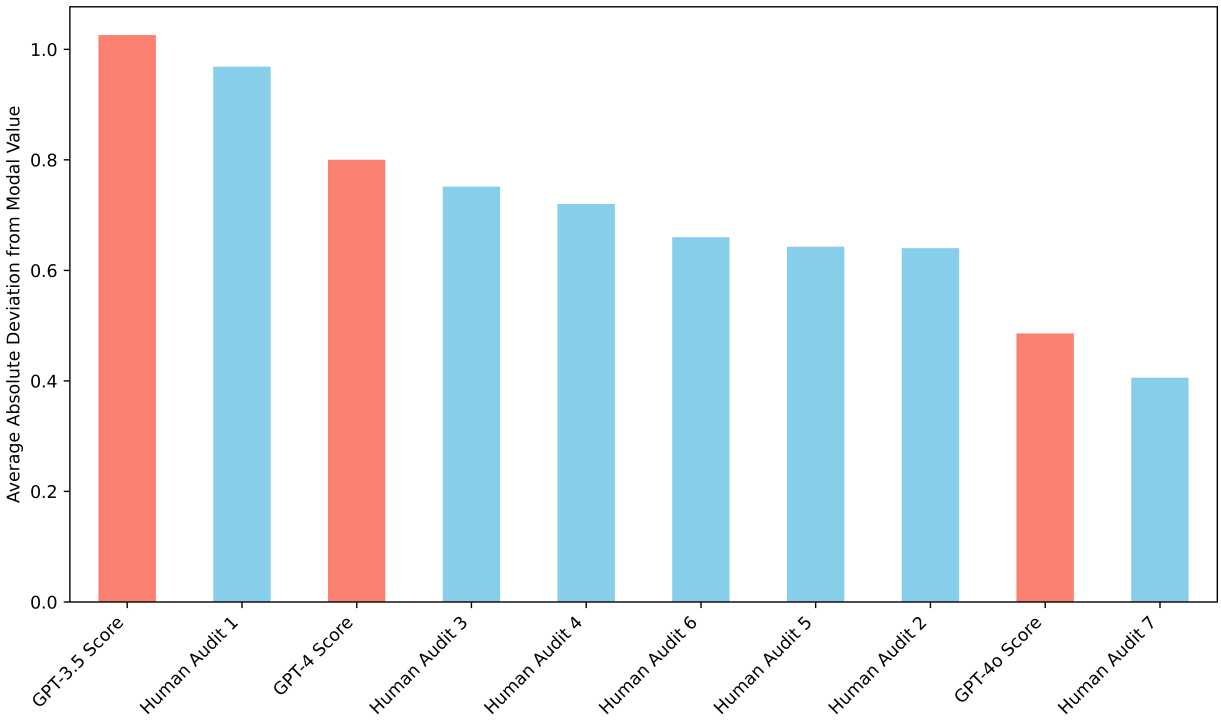
6.2 Human Audit

How reliable are the scores assigned by an LLM? Here, we conduct a human audit of the LLMs happiness scores. To this end, we first create a summary of each life history, extracting the aspects that were important for a person's happiness or misery. This is necessary to obtain texts of manageable length for the human auditors. The *summaries* are then scored by humans (while the LLMs assess the full text). Despite the extra summation step, we show how similar the human and LLM scores for retrospective happiness are.

We created summaries for 119 life narratives. We then used these summaries to obtain human coder assessments, employing a set of 7 research assistants (who had not previously seen the life narratives) on a subset of 50 summaries to score them on a scale of 1 to 7, based on a close reading of the summaries. Each text was assigned to seven readers. Humans largely agreed with each other's scores – the results of different humans deviated on occasion from the humans' modal value; LLMs also agree with the human evaluation (and each other) to a large extent. For a detailed evaluation of similarity for a range of models, see Appendix A.4. If we use correlations as a measure of similarity, we find that that can be as high as 0.84 between the average human score and GPT-4o (whereas the greatest level of inter-human agreement was 0.9).

Here, we evaluate the consistency between human auditors and three different generations of OpenAI LLMs (GPT-3.5, GPT-4, GPT-4o) using the modal human score, as in our previous validation exercises. Figure 11 presents a bar chart for each LLM and human scores, depicting agreement with the modal human score. The result is displayed in . Our human raters had absolute deviations from the modal human score ranging from 0.41 to 0.97. GPT-4o achieved the second-lowest av-

Figure 11: Average Absolute Deviation from Modal Answer for Different Human Evaluators and LLMs



Note: We show the agreement rate with the modal response of humans for seven research assistants as well as three generative models (GPT-3.5, 4, and 4o).

erage deviation, having a slightly higher difference from modal scores than RA 7 only. The high performance is not a foregone conclusion; GPT-3.5 showed the worst performance of any rater, with an average absolute deviation of over 1, and GPT-4 was not much better, with a deviation of 0.8, making it the second-lowest performer of all "raters". GPT-4o, on the other hand, comfortably outperforms most humans in terms of agreement with the modal human assessment.

6.3 Correlates of Happiness

Given the high rate of agreement between human and LLM scores, we extended the LLM scoring to the whole set of 1,335 life histories. Table 9 summarizes patterns in mean retrospective happiness scores by population sub-groups. Women and men are equally satisfied with their lives on average, and non-white narrators have 0.16 points lower average scores (about 0.16 standard deviations) than whites. Those in high-SES occupations are 0.26 points happier on average than workers in low-SES jobs, consistent with other evaluations of life (Kahneman and Deaton, 2010)

or measures of subjective well-being (see e.g. Stevenson and Wolfers, 2008, who find a strong correlation between individual income and Gallup's life satisfaction measures). Urban and rural differences are insignificant, as are the gaps between immigrants and the native-born.

Perhaps surprisingly, one large gap in average happiness is between those who migrated internally and those who did not. We find that this gap is similar in size when looking only at movers west, and only at movers to urban areas. Nor is it a mechanical function of document length, where longer documents are more likely to describe a move and more likely to convey a sense of happiness: movers are happier when restricting attention only to sufficiently long documents. For example, A.J. Manning, a real estate agent in Miami during the 1920s, took a trip to Florida in 1911. He recalls being "entranced with the beauty of the scenery" and relocated to Miami, building a business there. Similarly, African-Americans who moved out of the South often experienced considerable increases in their working conditions and educational prospects. Fred Dixon, reflecting on the reasons why blacks like him moved from Brewton, Alabama to Omaha, observed that "the packing industry was affording ... an opportunity to earn a higher wage ... Omaha also afforded a better educational opportunity for ... children which was upper-most in [movers'] minds."

Happiness scores were generally higher for those who found meaning in work relative to those indicating that it was merely "important". Agnes Harrell of Marion, S.C., was abandoned by her husband with 3 young children. She described with evident pride how she kept her family fed, clothed, and cared for:

> "days i'm at home there's three meals a day to cook and ... housework and the washing ... there's no time to recreate. even sit up here at night and make quilt squares. i know i've got the will power about me all right - know i've got a share to do long as i live. i believe when a woman lives seventeen years and keeps enough to eat for herself and three growing children does all the housework keeps the beds repaired and something for the youngsters to wear she just done well."

Nonetheless, it is clear that the work she does (taking in washing) is instrumental, not a source of pride or fulfillment beyond the money it provided. In contrast, others found a deep sense of satisfaction in their work, like the nurse Virginia Suffolk said: "I did grow up filled with a strong desire to be of some help to my people. After I completed my literary education, I finally persuaded my father to let me go in training for a nurse."[18]

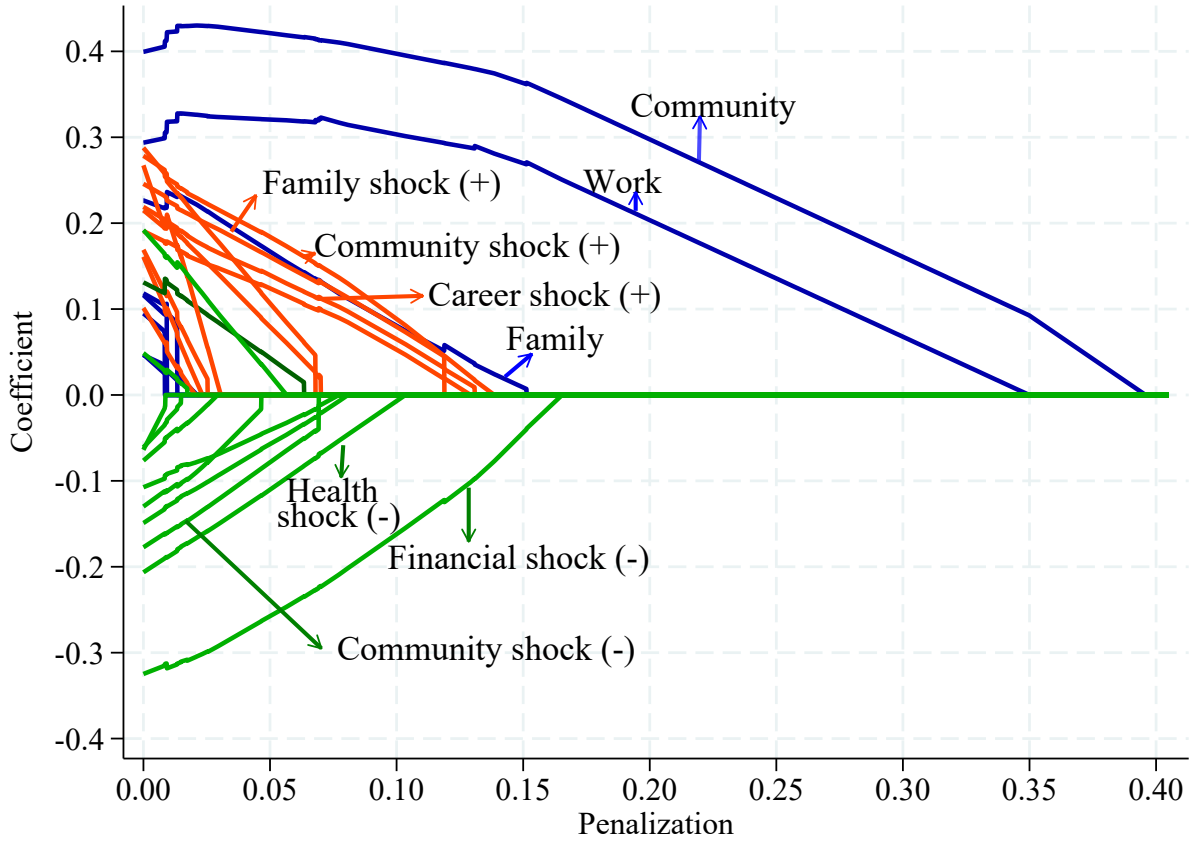
[18] Along similar lines, librarian Janie Smith, said that "[s]o completely happy am I in my work... I can ask for nothing better in life. ... I find so much happiness in my work, that I really need very little outside pleasure." Harrell, Smith and Harrell scored 5, 4 and 5 respectively on the happiness scale of 1-7.

Table 9: Comparing Retrospective Life Satisfaction in Different Sub-Groups

Variable	Category	Average	Difference (p-value)
Gender	Female	4.24	-0.08
			(0.06)
Race	Non-White	4.18	-0.14^{**}
			(0.08)
SES	High	4.45	0.31^{***}
			(0.05)
Urban Residency	Urban	4.27	-0.06
			(0.06)
Mover	Mover	4.48	0.38^{***}
			(0.05)
Immigration	Immigrant	4.26	-0.05
			(0.09)
Region, where an individual spent most of her life	Midwest	4.32	-0.03
			(0.08)
	Northeast	3.88	-0.49^{***}
			(0.08)
	South	4.21	-0.13^{**}
			(0.06)
	Southwest	4.54	0.36^{***}
			(0.06)
	West	4.42	0.13
			(0.13)

Note: The table shows the average life satisfaction in each sub-group, along with the difference from the rest (indicated by "Difference") and its significance level. Standard errors are in parentheses. Stars represent the p-value thresholds, where *** means 1%, ** means 5%, and * means 10%.

Figure 12: LASSO Predictors of Satisfaction



6.4 Statistical Predictors of Satisfaction

With the narrator's sources of life satisfaction, critical junctures, and retrospective satisfaction at hand, this section asks which elements of one's life trajectory predict a positive, retrospective evaluation of life during one's golden years.

The reconstructed memories quantified in Sections 4 and 5 to characterize one's lifetime trajectory are arguably measured with noise and are correlated with each other. We thus apply LASSO, a machine learning method that via regularization selects the variables (shrinkage) most predictive of satisfaction scores (Tibshirani, 1996).

Before presenting the results, we should stress that LASSO does not identify causal effects nor does it provide unbiased estimates of linear regression coefficients. The objective is to offer a parsimonious representation of what dimensions best predict a fulfilled life in a statistical sense. We compute the LASSO path using Least Angle Regression (LAR, Efron et al. (2004)), allowing λ to range from 0 (OLS) to infinity (all coefficients are set to zero). Figure 12 displays the

entire LASSO path for life satisfaction exploring the role of sources of life satisfaction and critical junctures. On the vertical axis, we plot the coefficient estimates. The horizontal axis shows the penalization parameter ranging from zero/no penalization (OLS), to regularization results in all correlates being zero. Moving from left to right, more variables drop from the model; the last variables that drop out are the "strongest" predictors of satisfaction.[19].

A point of clarification before we describe the patterns is due. Our measure of retrospective life satisfaction is not based on our narrators independently answering themselves how satisfied they are, but is based on how our readers (and LLMs) having read the life histories assess the level of positive life evaluation by narrators. Hence, the associations recovered below should be interpreted as answering the following question: What aspects of someone's life history predict how an objective reader determines how happy the life of the narrator has been?

While the number of observations is quite small for "machine learning" type algorithms, LASSO is useful in pinpointing the relatively more important predictors of life satisfaction. First, deriving meaning from nature and adventure and religion are variables with the least predictive power. Second, having found meaning in work and community are the top two most important predictors of happiness. Third, critical events that defined the narrators' economic fortunes are salient correlates of satisfaction. Financial shocks predict lower retrospective contentment whereas life events when career and family fortunes improved predict consistently a more satisfied narrator. Fourth, health shocks mostly matter as a negative and come with substantial costs in terms of overall life satisfaction.

The strong predictive power of having found meaning in one's career on overall life satisfaction coupled with how widespread is the former as a source of fulfillment, see Figure 7, highlights one of the novel findings of our American Life Histories. For many of our narrators work and career held comparable significance to close personal relationships challenging previous work that emphasizes the role of the latter as the primary source of life satisfaction.

7 Clustering Life Histories

In this section, we describe the joint patterns of critical junctures and sources of life meaning using clustering analysis. The goal is to help uncover the natural groupings in how the narrators

[19]We also computed the "optimal" degree of regularization through K-fold cross-validation. We split the data into 10 (5) equal size bins of narrators; for a given value of λ we fit (train) the model in 9 bins, then take out-of-sample fits in the remaining one, and compute the forecast accuracy with the mean square error criterion. Repeating this process for a grid of 100 values of λ, we take the value that yields the lowest error. The cross-validation routine yields λ equal to around 0.04. We interpret the estimates associated with this optimality result with caution given the discussion in Kolesár, Müller and Roelsgaard (2023)

described what made their lives worth living and which key events, or turning points, they felt worthy of inclusion in their life stories. Our approach divides the data into distinct groups using hierarchical agglomerative clustering, which has proven powerful in recent economic studies of life patterns (Humphries, 2022; Audoly et al., 2024). Intuitively, this method begins by treating each of the N individuals in the data as their own cluster, so that every person is their own unique "snowflake." The next layer of the hierarchy, with N-1 clusters, is created by combining the closest two clusters into one. The layer on top of that, with N-2 clusters, is formed by merging the next two closest clusters, and so forth, until there is only one cluster at the top of the hierarchy containing every observation.

In principle, we can report the results of this exercise for any number of clusters up to our number of observations. In practice, it is not useful to summarize a large number of clusters, like 20. We describe the data using six clusters, which is a parsimonious and still informative choice, similar to the four and seven clusters used by Audoly et al. (2024) and Humphries (2022) to summarize life-cycle patterns of wealth mobility and self-employment, respectively. We report how our results change when using alternative numbers of clusters below. We follow the studies cited above and use Ward's method for clustering, which tries to minimize variance within each cluster. We measure distance between two narratives using the Euclidean distance, based on each narrative's scores for sources of life meaning, presence of positive critical junctures, and presence of negative critical junctures. The scores for each source or critical juncture are the narratives' residuals from a logistic regression with controls for document length and writer fixed effects, consistent with our analysis above

To summarize the main distinguishing features of each cluster, we compute the odds ratio for each source of meaning or critical juncture, defined as the probability that the source is found in the cluster divided by the probability that the source is found in all other clusters. Table 10 summarizes the most salient features of the six clusters, defined as those for which the cluster has the highest odds ratio. The full set of odds ratios is available in Appendix Table 10. For exposition, we report up to the four most salient sources of meaning or critical junctures in Table 10. The percentage of all documents in each cluster is also reported.

In general, the clusters have sharp distinctions. Narratives in Cluster 1 draw relatively most meaning from community; their odds ratio of 2.5 means they are 2.5 times as likely to get happiness, satisfaction, meaning, or purpose from community as narratives in other clusters. Their most salient critical junctures are positive ones related to nature, community, and migration. An example narrative from Cluster 1 is Alcide Savoie, whose life turning points we described in Section 5. Born in Canada, he migrated to Barre, Vermont, where he joined the community and took a job as a stone worker like a significant number of other men in Barre at the time.

Table 10: Most Salient Sources of Meaning and Critical Junctures by Cluster

Cluster (percent)	Most Salient Sources of Meaning	Most Salient Critical Junctures
1 (12%)	Community (2.5)	(+) Nature (2.1) (+) Community (1.5) (+) Migration (1.4)
2 (10%)	Work (2.0)	(+) Health (4.2) (+) Personal (1.7) (+) Career (1.3)
3 (28%)	Family (4.2) Adventure (3.2) Nature (3.0)	(-) Personal (1.9) (+) Family (1.2) (-) Education (1.2)
4 (12%)	Education (55.3) Religion (4.4) Resilience (2.3)	(+) Education (2.7)
5 (28%)	[None]	(+) Historical (2.1) (-) Nature (1.9) (-) Community (1.7) (-) Historical (1.3)
6 (15%)	[None]	(-) Cultural (2.7) (+) Cultural (2.4) (-) Family (2.0) (-) Career (2.0)

Note: This table reports the most and least likely sources of happiness, satisfaction, and meaning for each cluster. The odds ratio for each source of happiness is computed as the probability that a topic is present for the cluster relative to the probability it is found in all other clusters. We report up to the four most salient categories for each cluster, computing as those having the maximum odds ratio across clusters.

Cluster 2 derives most meaning from work and has generally positive critical junctures related to health, personal growth, and career. An example from this cluster is the story of Willie Marlowe, which is one of occupational mobility. Willie was born into a family of poor tenant cotton farmers in South Carolina and left for the nearby city of Marion, where he eventually landed a "good job" as a baker where he worked long hours and earned significantly more than he did growing cotton.

Cluster 3 is the largest, making up 28 percent of all documents, and contains narratives that are

four times more likely than others to get satisfaction from family and relationships. This group is also more likely to derive meaning from past adventures and outdoor activities. Their most salient critical junctures are a mix of negative ones related to personal growth and education and positive ones about family. The story of Jeff Waggoner, which we summarized in Section 2, is part of this cluster. He describes his marriage to his wife Mary as "the wisest thing [I] ever did," and wistfully recounts his adventures living on the range with herds of cattle and his fellow cowboys.

Life histories in Cluster 4 are far and away the most likely to describe being uplifted by education, with an odds ratio of 55.3. They are also significantly more likely than others to derive meaning from resilience and religion. The only salient critical junctures in this cluster are positive ones related to education. An example in this cluster is Martha Ellen Devan, who was born in South Carolina as the youngest of ten children. Her parents owned a general store and could afford to send her to secondary school, which she finished at age 18, getting married a year later. She and her husband eventually moved to Jacksonville, Florida, where they ran their own successful store and became active members of the Methodist community.

The last two clusters expressed few or no sources of happiness or satisfaction in life. These two clusters generally captured a mix of difficult lives full of negative critical junctures and somewhat less informative documents focused more on external events rather than revealing much about themselves. Many of the latter can be found in Cluster 5, constituting 28 percent of all narratives, whose most common critical junctures were related to historical events – both positive and negative – and negative events related to nature or their communities. The story of Frank Faith is one typical example. Born in Wisconsin, he migrated to Nebraska and worked as a farmer and rancher for three decades before switching to carpentry work. His narrative discusses local horse thieves who were caught and hanged, the time he was robbed in a stagecoach, and an incident where a land squatter was run out of town by a posse of neighbors. He sends few signals about what brought him satisfaction, and his interviewer notes that he "probably lived a pretty rough life."

Cluster 6, representing 15 percent of all narratives, is heavily focused on negative experiences or shocks. Its most common critical junctures were cultural, both positive and negative, and members were twice as likely as other clusters to report a negative juncture related to career or family. Negative financial shocks were also most salient in this group. Minnie Marshall falls squarely into Cluster 6. Losing her mother when young and entering into an abusive marriage were both negative family-related shocks. Her continued difficulties securing stable work were another negative juncture. In general, Minnie gave little indication of anything positive in her life, and her interviewer describes Minnie's story as a "futile struggle of life from past to present."

Table 11 reports demographic and economic characteristics of narratives that were not used to form the clusters themselves. Any differences in these characteristics across clusters thus reveal

Table 11: Prevalence of Characteristics by Cluster

	Overall	Cluster Average					
Characteristic	Average	1	2	3	4	5	6
Female (%)	34	27	23	39***	43***	24	44***
Non-white (%)	14	10	14	14	23***	11	12
Urban (%)	45	52**	44	41	56***	44	41
Immigrant (%)	12	15	13	12	7	8	15*
Migrant (%)	54	56	59*	58	64***	54	47
Occupation SEI	29	28	29	27	42***	30	19
Happiness	4.8***	4.6***	4.6***	4.7***	5.1***	3.8	3.6
Agreeable (%)	62	63	65	81***	85***	40	43

Notes: This table reports the prevalence of characteristics of the narratives by cluster, plus the average Duncan Occupation SEI and Happiness score. Stars represent the *p*-values of the tests that the value is higher than average, where *** means 1%, ** means 5%, and * means 10%.

independent features that underlie the clusters rather than the clustering procedure itself. Female narrators are significantly more likely to appear in Clusters 3 (emphasizing family), 4 (education), and 6 (negative shocks). Men are most concentrated in Cluster 2 (focused on work) and Cluster 5 (historical events and negative junctures). These patterns highlight how gender is central in shaping both the sources of positivity in uplifting life stories and the determinants of negativity in more difficult lives. Among the least positive stories, women focus relatively more on individual setbacks and challenges, while men concentrate more on broader events happening in their communities.

Non-whites are most prevalent in Cluster 4, appearing about twice as frequently there as in Cluster 1 or either of the negative clusters.[20] Religion is a much more common source of meaning in the lives of non-whites, who often express gratitude for learning opportunities, whether in formal education or trades, for their families or for themselves. Urban narrators are more prevalent in Cluster 1 (community) and Cluster 4, with the latter finding consistent with greater educational opportunities in cities than in the countryside. Immigrants fall roughly evenly between the different groups, with no statistically significant differences between them. Internal migrants and those with higher occupational scores are more likely to be in Cluster 4, likely reflecting the educational opportunities they acknowledged in their life histories.

[20] In Appendix A.2, we offer a potential explanation for this finding highlighting the differential migration destination between whites and non-whites in late 19th and early 20th-century US.

Despite their differences, retrospective life happiness scores are very similar across the four positive clusters, averaging around 5. Contrary to Tolstoy's suggestion in *Anna Karenina* that all happy families are alike, these narratives reveal multiple paths to a life looked back upon with satisfaction. Life histories in the last two clusters score about a full point lower, a difference that is statistically significant at the one-percent level. The highest agreeableness scores appear in Clusters 3 and 4, while the lowest appear in the two negative clusters.

The upshot of this analysis is that clustering the life histories based on their narrators' sources of meaning and critical life junctures results in groups with sharp and meaningful distinctions along characteristics not directly targeted. This finding echoes the cleavages across population sub-groups we described in Sections 4 and 5, and also helps illustrate the common features that connect key life experiences to what makes life worth living.[21]

8 Conclusion

> "[t]he premise of essentially all economics . . . is that leisure is good and work is bad. . . . economics is going to have to find a way to recognize the fundamental human satisfactions that come from making a contribution . . ."
>
> Larry Summers

Numerous studies in economics have investigated what makes people "happy" (Frey, 2010; Di Tella and MacCulloch, 2006). Psychologists distinguish happiness from life satisfaction, emphasizing the fleeting nature of the former ("pursuit of pleasure") and the slow-moving and more profound aspects of the latter (Peterson, Park and Seligman, 2005). We take a unique perspective, examining what older respondents highlight as key aspects of their lives as they look back. To this end, we empirically analyze what brought meaning and a sense of fulfillment to the lives of 1,300 Americans interviewed during the 1930s. These life narratives were collected as part of a federal make-work program, the Federal Writers Project. Though the respondents generally lived unremarkable lives by objective measures, nearly all described rich experiences filled with ups and

[21] The choice of six clusters is parsimonious but informative, as the tables above indicate. Yet other numbers of clusters may also be informative. Appendix Figure A17 illustrates how our six clusters are agglomerated up from ten clusters – and combined down to just one – using a dendrogram.

downs, and most reported satisfaction with their lives in retrospect.

We use frontier LLM models to extract factual information and complex assessments of people's lives and views from the recorded interviews. Extensive human audit testing shows that under the right circumstances, the LLM can be every bit as reliable as human readers. With this result in hand, we gather information on what brought meaning and purpose to people's lives. Some of our findings demonstrate the veracity of folk wisdom – women care more about family and relationships than men. Others offer new insights: In addition to the importance of relationships and family, we find that the roles of work and of contributing to communities and society loom large. Men and women reminisce proudly about how they learned their trades, emphasizing their mastery of complex tasks. Such pride is by no means limited to the professionals in our dataset; they emphasize how their work allows them to leverage their skills like the librarian putting their love of learning to good use. Many respondents eagerly share their insights into mundane businesses, from the turpentine industry to bone-collecting after droughts. Work in these life narratives is not necessarily drudgery, reluctantly endured; in most cases, it is a source of pride, joy, social recognition, and standing in the community. In other words, many of the men and women in the American Life Histories corpus are *homo faber* ("man the maker" in Latin) – creatures that need to work, produce, and contribute to flourish.

References

ADUKIA, A., AND E. HARRISON (2024): "Separation of Church and State Curricula? Examining Public and Religious Private School Textbooks," Unpublished Working Paper, University of Chicago.

AKERLOF, G. A., AND R. E. KRANTON (2000): "Economics and identity," *The quarterly journal of economics*, 115(3), 715–753.

ALMELHEM, A., M. IYIGUN, A. KENNEDY, AND J. RUBIN (2023): "Enlightenment Ideals and Belief in Progress in the Run-up to the Industrial Revolution: A Textual Analysis," IZA Discussion Papers 16674, Institute of Labor Economics (IZA).

AMIRIZANIANI, M., E. MARTIN, M. SIVACHENKO, A. MASHHADI, AND C. SHAH (2024): "Do LLMs Exhibit Human-Like Reasoning? Evaluating Theory of Mind in LLMs for Open-Ended Responses," *arXiv preprint arXiv:2406.05659*.

ANDERSEN, L., AND J. BENTZEN (2022): "In the Name of God! Religiosity and the Transition to Modern Growth," Working Paper 16938, CEPR Press.

ANDRE, P., I. HAALAND, C. ROTH, M. WIEDERHOLT, AND J. WOHLFART (2024): "Narratives about the Macroeconomy," Unpublished Working Paper, Goethe University Frankfurt.

ANTONOVSKY, A. (1993): "The structure and properties of the sense of coherence scale," *Social science & medicine*, 36(6), 725–733.

ARIELY, D., E. KAMENICA, AND D. PRELEC (2008): "Man's Search for Meaning: The Case of Legos," *Journal of Economic Behavior and Organization*, 67, 671–677.

ASH, E., AND S. HANSEN (2023): "Text algorithms in economics," *Annual Review of Economics*, 15(1), 659–688.

ASHRAF, N., O. BANDIERA, AND B. K. JACK (2014): "No Margin, No Mission? A Field Experiment on Incentives for Public Services Delivery," *Journal of Public Economics*, 120, 1–17.

ASHRAF, N., O. BANDIERA, V. MINNI, AND L. ZINGALES (2024): "Meaning at Work," Unpublished Working Paper, London School of Economics.

ASHWIN, J., A. CHHABRA, AND V. RAO (2023): "Using Large Language Models for Qualitative Analysis can Introduce Serious Bias," Policy Research Working Paper 10597.

AUDOLY, R., R. MCGEE, S. OCAMPO, AND G. PAZ-PARDO (2024): "The Life-Cycle Dynamics of Wealth Mobility," Unpublished Working Paper, University of Western Ontario.

AYYAR, S., U. BOLT, E. FRENCH, AND C. O'DEA (2024): "Imagine your Life at 25: Gender Conformity and Later-Life Outcomes," Working Paper 32789, National Bureau of Economic Research.

BAKER, S. R., N. BLOOM, AND S. J. DAVIS (2016): "Measuring economic policy uncertainty," *The quarterly journal of economics*, 131(4), 1593–1636.

BANKS, A. (1980): *First-Person America*. Alfred A. Knopf.

BARTIK, A. W., A. GUPTA, AND D. MILO (2024): "The Costs of Housing Regulation: Evidence From Generative Regulatory Measurement," Unpublished Working Paper, University of Illinois.

BAZZI, S., M. FISZBEIN, AND M. GEBRESILASSE (2020): "Frontier culture: The roots and persistence of "rugged individualism" in the United States," *Econometrica*, 88(6), 2329–2368.

BEACH, B., AND W. W. HANLON (2023): "Historical Newspaper Data: A Researcher's Guide," *Explorations in Economic History*, 90, 101541.

BENABOU, R., A. FALK, AND J. TIROLE (2020): "Narratives, Imperatives, and Moral Reasoning," *Unpublished Working Paper, Princeton University*.

BESLEY, T., AND M. GHATAK (2005): "Competition and Incentives with Motivated Agents," *The American Economic Review*, 95(3), 616–636.

BESLEY, T., AND T. PERSSON (2023): "The political economics of green transitions," *The Quarterly Journal of Economics*, 138(3), 1863–1906.

BICK, A., N. FUCHS-SCHÜNDELN, AND D. LAGAKOS (2018): "How do Hours Worked Vary with Income? Cross-Country Evidence and Implications," *American Economic Review*, 108(1), 170–199.

BORCHERT, S. (2021): *Republic of Detours: How the New Deal Paid Broke Writers to Rediscover America*. Farrar, Straus and Giroux.

BORDALO, P., G. BURRO, K. COFFMAN, N. GENNAIOLI, AND A. SHLEIFER (Forthcoming): "Imagining the Future: Memory, Simulation, and Beliefs," *Review of Economic Studies*.

BORDALO, P., N. GENNAIOLI, AND A. SHLEIFER (2020): "Memory, Attention, and Choice," *Quarterly Journal of Economics*, 135(3), 1399–1442.

BOTKIN, B. A. (1945): *Lay My Burden Down: A Folk History of Slavery*. University of Chicago Press.

CASSAR, L., AND S. MEIER (2018): "Nonmonetary Incentives and the Implications of Work as a Source of Meaning," *Journal of Economic Perspectives*, 32(3), 215–238.

COSTA, D. L. (2000): "The Wage and the Length of the Work Day: From the 1890s to 1991," *Journal of Labor Economics*, 18(1), 156–181.

CRAWFORD, S. C. (1980): *Quantified Memory: a Study of the WPA and Fisk University Slave Narrative Collections*. The University of Chicago.

DAL BÓ, E., F. FINAN, AND M. A. ROSSI (2013): "Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service," *Quarterly Journal of Economics*, 128(3), 1169–1218.

DELL, M. (Forthcoming): "Deep Learning for Economists," *Journal of Economic Literature*.

DELL, M., J. CARLSON, T. BRYAN, E. SILCOCK, A. ARORA, Z. SHEN, L. D'AMICO-WONG, Q. LE, P. QUERUBIN, AND L. HELDRING (2023): "American Stories: A Large-Scale Structured Text Dataset of Historical US Newspapers," *Advances in Neural Information and Processing Systems, Datasets and Benchmarks*.

DI TELLA, R., AND R. MACCULLOCH (2006): "Some Uses of Happiness Data in Economics," *Journal of Economic Perspectives*, 20(1), 25–46.

ENKE, B. (2024): "Moral boundaries," *Annual Review of Economics*, 16(1), 133–157.

ESCOTT, P. D. (1979): *Slavery Remembered: A Record of Twentieth-Century Slave Narratives*. The University of North Carolina Press, Chapel Hill.

FISHBACK, P. V. (1998): "Operations of "Unfettered" Labor Markets: Exit and Voice in American Labor Markets at the Turn of the Century," *Journal of Economic Literature*, 36(2), 722–765.

FITOUSSI, J.-P., J. E. STIGLITZ, AND A. K. SEN (2011): *Mismeasuring our lives: why GDP doesn't add up*. ReadHowYouWant. com.

FLYNN, J. P., AND K. A. SASTRY (2024): "The Macroeconomics of Narratives," Unpublished Working Paper, Princeton University.

FRANKL, V. E. (1963): *Man's search for meaning*. Simon and Schuster.

FREY, B. S. (2010): *Happiness: A Revolution in Economics*. MIT press.

GEIECKE, F., AND X. JARAVEL (2024): "Conversations at Scale: Robust AI-led Interviews with a Simple Open-Source Platform," CEPR Discussion Paper No. 19705.

GENTZKOW, M., B. T. KELLY, AND M. TADDY (2019): "Text as Data," *Journal of Economic Literature*, 57, 535–574.

GILARDI, F., M. ALIZADEH, AND M. KUBLI (2023): "ChatGPT Outperforms Crowd Workers for Text-Annotation Tasks," *Proceedings of the National Academy of Sciences*, 120(30).

GRISWOLD, W. (2016): *American Guides: The Federal Writers' Project and the Casting of American Culture*. University of Chicago Press.

HANDLAN, A. (2024): "Text Shocks and Monetary Surprises: Text Analysis of FOMC Statements with Machine Learning," Unpublished Working Paper, Brown University.

HARRINGTON, F., F. S. KERR, AND H. G. ALSBERG (1939): *These Are Our Lives*. The University of North Carolina Press.

HASSAN, T. A., S. HOLLANDER, A. KALYANI, L. VAN LENT, M. SCHWEDELER, AND A. TAHOUN (Forthcoming): "Economic Surveillance using Corporate Text," *Journal of Economic Perspectives*.

HEINTZELMAN, S. J., AND L. A. KING (2014): "Life is pretty meaningful.," *American psychologist*, 69(6), 561.

HIRSCH, J. (2003): *Portrait of America: A Cultural History of the Federal Writers' Project*. University of North Carolina.

HUMPHRIES, J. E. (2022): "The Causes and Consequences of Self-Employment over the Life Cycle," Unpublished Working Paper, Yale University.

JONES, C. I., AND P. J. KLENOW (2016): "Beyond GDP? Welfare Across Countries and Time," *American Economic Review*, 106(9), 2426–2457.

KAHANA, M. (2012): *Foundations of Human Memory*. Oxford University Press.

KAHNEMAN, D., AND A. DEATON (2010): "High income improves evaluation of life but not emotional well-being," *Proceedings of the national academy of sciences*, 107(38), 16489–16493.

KALYANI, A. (2024): "The Creativity Decline: Evidence from U.S. Patenting," Unpublished Working Paper, Federal Reserve Bank of St. Louis.

KALYANI, A., N. BLOOM, M. CARVALHO, T. A. HASSAN, J. LERNER, AND A. TAHOUN (Forthcoming): "The Diffusion of New Technologies," *Quarterly Journal of Economics*.

KAPLAN, G., AND S. SCHULHOFER-WOHL (2018): "The Changing (Dis-)Utility of Work," *Journal of Economic Perspectives*, 32(3), 239–258.

KAYAM, O. (2018): "The Readability and Simplicity of Donald Trump's Language," *Political Studies Review*, 16(1), 73–88.

KING, L. A., AND J. A. HICKS (2021): "The science of meaning in life," *Annual review of psychology*, 72(1), 561–584.

KISSINGER, H. A. (1950): *The Meaning of History (Reflections on Spengler, Toynbee and Kant)*. Bokförlaget Stolpe, Stockholm.

KNUDSEN, A. S. B. (2019): "Those Who Stayed: Individualism, Self-Selection and Cultural Change during the Age of Mass Migration," Discussion papers, University of Copenhagen. Department of Economics.

KOLESÁR, M., U. K. MÜLLER, AND S. T. ROELSGAARD (2023): "The Fragility of Sparsity," *arXiv preprint arXiv:2311.02299*.

KORINEK, A. (2023): "Generative AI for Economic Research: Use Cases and Implications for Economists," *Journal of Economic Literature*, 61(4), 1281–1317.

LUDWIG, J., AND S. MULLAINATHAN (2024): "Machine Learning as a Tool for Hypothesis Generation," *Quarterly Journal of Economics*.

LUDWIG, J., S. MULLAINATHAN, AND A. RAMBACHAN (2025): "Large Language Models: An Applied Econometric Framework," Working Paper 33344, National Bureau of Economic Research.

MA, Z., J. SANSOM, R. PENG, AND J. CHAI (2023): "Towards a holistic landscape of situated theory of mind in large language models," *arXiv preprint arXiv:2310.19619*.

MAIDEU-MORERA, G. (2024): "Amenity-Biased Technical Change," Unpublished Working Paper, Toulouse School of Economics.

MAIRESSE, F., M. A. WALKER, M. R. MEHL, AND R. K. MOORE (2007): "Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text," *Journal of Artificial Intelligence Research*, 30, 457–500.

MANGIONE, J. (1996): *The Dream and the Deal: The Federal Writers' Project, 1935-1943*. Syracuse University Press.

MCADAMS, D. P. (2001): "The Psychology of Life Stories," *Review of General Psychology*, 5(2).

MICHALOPOULOS, S., AND C. RAUH (2024): "Movies," Working Paper 32220, National Bureau of Economic Research.

MICHALOPOULOS, S., AND M. XUE (2021): "Folklore," *Quarterly Journal of Economics*, 136(4), 1993–2046.

OISHI, S., AND E. DIENER (2014): "Residents of poor nations have a greater sense of meaning in life than residents of wealthy nations," *Psychological Science*, 25(2), 422–430.

PENKOWER, M. N. (1977): *The Federal Writers' Project*. University of Illinois Press.

PENNEBAKER, J. W., M. R. MEHL, AND K. G. NIEDERHOFFER (2003): "Psychological Aspects of Natural Language Use: Our Words, Our Selves," *Annual Review of Psychology*, 54(1), 547–577.

PETERSON, C., N. PARK, AND M. E. SELIGMAN (2005): "Orientations to Happiness and Life Satisfaction: The Full Life Versus the Empty Life," *Journal of Happiness Studies*, 6, 25–41.

RUGGLES, S., M. A. NELSON, M. SOBEK, C. A. FITCH, R. GOEKEN, J. D. HACKER, E. ROBERTS, AND J. R. WARREN (2024): "IPUMS Ancestry Full Count Data: Version 4.0 [dataset]," *Minneapolis, MN: IPUMS*.

SCHMITT, D. P., M. VORACEK, A. REALO, AND J. ALLIK (2008): "Why Can't a Man Be More Like a Woman? Sex Differences in Big Five Personality Traits Across 55 Cultures," *Journal of Personality and Social Psychology*, 94(1), 168–182.

SHILLER, R. J. (2017): "Narrative Economics," *American Economic Review*, 107, 967–1004.

SHOPES, L. (2002): "Oral History and the Study of Communities: Problems, Paradoxes, and Possibilities," *The Journal of American History*, 89(2), 588–598.

SINGH, A., A. EHTESHAM, S. KUMAR, AND T. T. KHOEI (2024): "Enhancing AI Systems with Agentic Workflows Patterns in Large Language Model," in *2024 IEEE World AI IoT Congress (AIIoT)*, pp. 527–532. IEEE.

STANTCHEVA, S. (2021): "Understanding tax policy: How do people reason?," *The Quarterly Journal of Economics*, 136(4), 2309–2369.

STEGER, M. F. (2012): "Making meaning in life," *Psychological Inquiry*, 23(4), 381–385.

STEVENSON, B., AND J. WOLFERS (2008): "Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox," *Brookings Papers on Economic Activity*, 102(5), 1–87.

STRACHAN, J. W., D. ALBERGO, G. BORGHINI, O. PANSARDI, E. SCALITI, S. GUPTA, K. SAXENA, A. RUFO, S. PANZERI, G. MANZI, ET AL. (2024): "Testing theory of mind in large language models and humans," *Nature Human Behaviour*, pp. 1–11.

TAYLOR, D. A. (2009): *Soul of a People: The WPA Writers' Project Uncovers Depression America*. John Wiley Sons, Inc.

TERKEL, S. (1970): *Hard Times: An Oral History of the Great Depression*. Pantheon Books.

TIBSHIRANI, R. (1996): "Regression Shrinkage and Selection via the Lasso,," *Journal of the Royal Statistical Society. Series B (Methodological)*, 58, 267–288.

TURING, A. M. (1950): "Computing Machinery and Intelligence," *Mind*, 59(236), 433–460.

VAILLANT, G. E. (2008): *Aging Well: Surprising Guideposts to a Happier Life from the Landmark Study of Adult Development*. Hachette UK.

——— (2012): *Triumphs of Experience: The Men of the Harvard Grant Study*. Harvard University Press.

ZHONG, R., P. ZHANG, S. LI, J. AHN, D. KLEIN, AND J. STEINHARDT (2023): "Goal Driven Discovery of Distributional Differences via Language Descriptions," *Arxiv*.

ZHOU, P., A. ZHU, J. HU, J. PUJARA, X. REN, C. CALLISON-BURCH, Y. CHOI, AND P. AMMANABROLU (2022): "I cast detect thoughts: Learning to converse and guide with intents and theory-of-mind in dungeons and dragons," *arXiv preprint arXiv:2212.10060*.

Online Appendices

A Appendix

A.1 GPT Prompts

GPT Prompt for Representative Higher-Level Characteristics.

You will be provided with text delimited by triple backticks. Your task is to decide whether *there is a mention of the death of their own father or mother in the text.* Give your answer as a yes or a no. If yes, give evidence of the death of their parent using lines from the text in a JSON object with ”answer” and ”evidence” as keys. If no, format your response as a JSON object with ”answer” and ”evidence” as keys and set evidence as ”unknown.””

The italicized portion of the question varied from question to question. For parental loss when young we asked whether *the person mentions that their father died when the person was young.* For poverty kept me out of school we asked whether *this person indicates that poverty, or lack of resources, kept them out of school.* For religion we asked whether *religion is important to them.* For agreeableness we asked whether *this person has an agreeable personality.* For showing empathy we asked whether *this person shows empathy towards other people.*

Dictionary Method. In Section 3 we compared human readings to those of GPT and a dictionary method. The exact description of our dictionary method is as follows. For death of a parent, we looked for the mention of synonyms of “father” or “mother” or “parent” and “death” within 15 words of each other. The synonyms for father are ”dad”, ”daddy”, ”father”, ”pa”, ”papa”, ”pappa”, ”pappy”, ”pop”, ”Dad”, ”Daddy”, ”Father”, ”Pa”, ”Papa”, ”Pappa”, ”Pappy”, and ”Pop.” The synonyms for mother are ”ma”, ”mama”, ”mamma”, ”mammy”, ”mom”, ”mommy”, ”mother”, ”Ma”, ”Mama”, ”Mamma”, ”Mammy”, ”Mom”, ”Mommy”, and ”Mother.” The synonyms for parent are ”parent”, ”parents”, ”Parent”, and ”Parents.” The synonyms for death are ”buried”, ”dead”, ”death”, ”deceased”, ”die”, ”died”, ”grave”, ”graves”, ”kill”, ”killed”, ”lived to be”, ”loss”, ”murder”, ”murdered”, ”passed away”, ”passed on”, ”succomb”, ”succombed”, ”succumb”, ”succumbed”, ”widowed”, ”Buried”, ”Dead”, ”Death”, ”Deceased”, ”Die”, ”Died”, ”Grave”, ”Graves”, ”Kill”, ”Killed”, ”Lived to be”, ”Loss”, ”Murder”, ”Murdered”, ”Passed away”, ”Passed on”, ”Succomb”, ”Succombed”, ”Succumb”, ”Succumbed”, and ”Widowed.”

For parental loss when young look additionally for the mention of one of these words within 15 words of the parental loss being mentioned in the text: ”1”, ”2”, ”3”, ”4”, ”5”, ”6”, ”7”, ”8”, ”9”,

”10”, ”11”, ”12”, ”13”, ”14”, ”15”, ”16”, ”17”, ”18”, ”19”, ”20”, ”one”, ”two”, ”three”, ”four”, ”five”, ”six”, ”seven”, ”eight”, ”nine”, ”ten”, ”eleven”, ”twelve”, ”thirteen”, ”fourteen”, ”fifteen”, ”sixteen”, ”seventeen”, ”eighteen”, ”nineteen”, ”twenty”, ”adolescent”, ”bloke”, ”boy”, ”boyhood”, ”chap”, ”child”, ”childhood”, ”formative”, ”girl”, ”girlhood”, ”infancy”, ”infant”, ”kid”, ”lad”, ”lass”, ”newborn”, ”schoolboy”, ”schoolgirl”, ”teen”, ”teenager”, ”tennage”, ”toddler”, ”young”, ”younger”, ”youth”, ”One”, ”Two”, ”Three”, ”Four”, ”Five”, ”Six”, ”Seven”, ”Eight”, ”Nine”, ”Ten”, ”Eleven”, ”Twelve”, ”Thirteen”, ”Fourteen”, ”Fifteen”, ”Sixteen”, ”Seventeen”, ”Eighteen”, ”Nineteen”, ”Twenty”, ”Adolescent”, ”Bloke”, ”Boy”, ”Boyhood”, ”Chap”, ”Child”, ”Childhood”, ”Formative”, ”Girl”, ”Girlhood”, ”Infancy”, ”Infant”, ”Kid”, ”Lad”, ”Lass”, ”Newborn”, ”Schoolboy”, ”Schoolgirl”, ”Teen”, ”Teenager”, ”Teenage”, ”Toddler”, ”Young”, ”Younger”, and ”Youth.” The numbers correspond to the mention of age in the context to capture for e.g., “My father passed away when I was four years old.”

For the dictionary approach to capturing an agreeable personality, hard work is important, or religion is important, we used the LIWC dictionaries. Agreeableness was based on the following categories: positive tone, positive emotion, Prosocial behavior, Politeness. Religion is important is based on LIWC dictionary for religion. Hard work is based on the LIWC dictionary associated with “work, drives and achievements.”

GPT-4o Simple Categorization Prompt

We aim to categorize a list of topics extracted from multiple biographies, which describe the sources of happiness, satisfaction, meaning, or purpose for the narrator, into broader ”classifications. You will be provided with a topic and reasoning for why the topic brings happiness to the reader. Your responsibility is to align this topic with just one of the following categories:

1. Family.
2. Work and Career.
3. Community and Service.
4. Adventure, Exploration, and Excitement.
5. Nature and Outdoor Activities.
6. Religious and Spiritual Beliefs.
7. Resilience and Overcoming Challenges.
8. Education.
9. Other.

While some topics may seem suitable for multiple categories, it is crucial to assign them to only one. Use the 'Other' category only as an exception when no other categories fit the topic. Use the provided reasoning for the topic to ensure the topic is actually about the chosen category because it is difficult to determine the appropriate category from the topic description alone. The result should be formatted with the topic, followed by a semicolon, and then the chosen category. For instance: 'Job in the mines; Work and Career.' It is essential to adhere to this format consistently.

Here is the topic and reasoning for classification:

Prompt for Happiness Score

Based on the biography provided, analyze the key events and overall tone to determine if the main character seemed to have lived a happy, content, and meaningful life.
Please score the overall happiness of this text on a scale from 1 to 7, using the following criteria:
Score 1 (Very Unhappy): The text is overwhelmingly negative, expressing deep sadness, anger, frustration, or pessimism, with an absence of positive sentiments or hope.
Score 2 (Unhappy): The text predominantly leans towards negative emotions, showing signs of unhappiness, dissatisfaction, or pessimism, with limited positive elements.
Score 3 (Mildly Unhappy): The text contains a mix of positive and negative sentiments, but leans slightly towards unhappiness or discontent, with some positive aspects present.
Score 4 (Neutral): The text balances positive and negative sentiments, expressing neither particularly happy nor unhappy emotions, or it might convey contentment without strong feelings either way.
Score 5 (Mildly Happy): The text contains a mix of positive and negative sentiments, but leans slightly towards happiness or contentment, with some negative aspects present.
Score 6 (Happy): The text predominantly leans towards positive emotions, demonstrating happiness, satisfaction, or optimism, with limited negative elements.
Score 7 (Very Happy): The text is overwhelmingly positive, expressing joy, gratitude, love, or excitement, with an absence of negative sentiments and an overall uplifting mood.
Begin your response with "Score: " followed by the number corresponding to your assessment. Then, provide a justification for the score, including specific examples and quotes from the text that support your evaluation.
Here is the text to analyze:

GPT-4o Topics Prompt with Guidelines

You will be provided with text delimited by triple backticks. Please identify topics that bring happiness, satisfaction, meaning, or purpose into the life of the narrator. For each topic, please answer in a JSON object with 'topic_name', 'reasoning', and 'evidence' as keys. 'topic_name' should be populated with a detailed description of the topic. 'reasoning' should include a short paragraph with your reasoning for the topic. 'evidence' should include a quote from the text that supports your reasoning.

Guidelines:

1. Incorporate all the pertinent topics you uncover that contributed to the narrator's happiness, satisfaction, meaning, or sense of purpose.

2. Please specify only those topics that are of great importance and have a significant impact on the narrator.

3. You should distinguish between sources of happiness and problems. When people talk about how they coped with their problems and complained, it is not a source of happiness. Please focus more on what the narrator is saying than what the interviewer, writer, or other characters in the text are saying.

4. Only include topics that have clear support from the text.

5. The evidence you choose must highlight the positive or uplifting nature of the topic. Simply mentioning the topic is not sufficient evidence. The evidence should portray the topic in a favorable light. Look for use of emotional language or clear statements of preference.

6. The same sentence cannot be used as evidence for two different topics.

7. The output should be formatted as a list of json objects, i.e.: [”topic_name”: ” ”, ”reasoning”: ” ”, ”evidence”: ” ”, ..., ”topic_name”: ” ”, ”reasoning”: ” ”, ”evidence”: ” ”, ...].

8. If you find nothing related to happiness, satisfaction, meaning or purpose in the life of the ”narrator then return a blank JSON, i.e. ” ”.

GPT-4o Enhanced Categorization Prompt

We aim to categorize a list of topics extracted from multiple biographies, which describe the sources of happiness, satisfaction, meaning, or purpose for the narrator, into broader classifications. You will be provided with a topic and reasoning for why the topic brings happiness to the reader. Your responsibility is to align this topic with just one of the following categories:

1. Family. If a person says that relationships with their family bring them happiness, then add a topic to this category.
2. Work and Career. Put a topic in this category if a person is talking about work, career, personal success, self-sufficiency/independence or achievements. Note that a wide range of topics can be assigned to the 'Work and Career', such as economic stability, entrepreneurship, trade, political career, and the work done by preachers.
3. Community and Service.
4. Adventure, Exploration, and Excitement.
5. Nature and Outdoor Activities.
6. Religious and Spiritual Beliefs.
7. Resilience and Overcoming Challenges.
8. Education.
9. Other.

While some topics may seem suitable for multiple categories, it is crucial to assign them to only one. Use the 'Other' category only as an exception when no other categories fit the topic. Use the provided reasoning for the topic to ensure the topic is actually about the chosen category because it is difficult to determine the appropriate category from the topic description alone.

The result should be formatted with the topic, followed by a semicolon, and then the chosen category. For instance: 'Job in the mines; Work and Career.' It is essential to adhere to this format consistently. Here is the topic and reasoning for classification:"

A.2 Movers: Timing and Destination Differences By Race

In this section, we shed some light on the seemingly puzzling finding that non-whites are more likely to find meaning in education in their lives. First, it is important to recognize that this difference is salient only among movers. In Figure A14, we break down the sources of the meaning of life by distinguishing between mover and stayer by race. What is clear is that white internal migrants, compared to non-white movers, are 18% percentage points less likely to mention education. Among non-movers, the racial gap in education as a source of life meaning is statistically insignificant. Why is this the case? This pattern may be partially explained by the fact that among our narrators, non-whites are systematically more likely to have moved to the urbanizing centers out of the South whereas whites consist of both those moving to the frontier during the 19th century as well as the urbanizing hubs during the same time period.

Specifically, the median year of the first move for the whites is 1884 and 1898 for the non-whites; see Figure A15a. For whites, this mainly reflects the movements that took place towards frontier regions, whereas for the non-whites, the years of move distribution maps into the end of Reconstruction (1877) and the Great Migration of African Americans starting in 1910 from the rural South to urban areas in the North, Midwest, and West of the United States. This racial asymmetry in destination migration is also reflected in the fact that 58 out of 91 non-whites (64%) and 277 out of 612 (45%) whites moved to areas that were more industrialized than their origins; see Figure A15b. We estimate the latter by focusing on each individual's first relocation and construct the change in the employment share of the 7-year average of overall manufacturing in the destination minus the corresponding average for the origin the last 7 years before the move. If this quantity is positive, it suggests that the individual moved into a place that was more industrialized than her origin and vice versa.

This asymmetry in the migration patterns may help explain why non-white movers appear to have found meaning in education more often than whites. Simply put, the pioneers, compared to those who moved to the manufacturing centers of the North, were less likely to find meaning in education.

A.3 The Geography of Meaning of Life and Critical Junctures

It is instructive to explore the geography of the meaning of life and the critical junctures. We distinguish between the 5 broad US regions.[22] Figure A10 tabulates where respondents from

[22]We split the contiguous US states into five regions: Northeast comprises Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, New York, New Jersey, and Pennsylvania. Midwest consists of Ohio, Michi-

different regions found meaning in their lives. The starkest regional gap is in terms of family as a source of life satisfaction with 82% of the narrators in the South finding satisfaction in the latter. This is 28 percentage points ahead of the Northeast, the second highest across regions, and a staggering 47 percentage points gap compared to those in the Midwest. In contrast, adventure in the South is the least common source, with those in the West and Southwest ranking first and second, respectively.

Except for adventure and the "other" category, the South stands out, suggesting that individuals in the South, compared to the rest of the country, derived meaning from a multitude of sources. This pattern may seem puzzling in the context of the South being less happy than the rest of the US. This may be resolved by looking at the intensity of positive and negative critical junctures Southerners report. Figure A11 reveals that deriving life meaning from family relationships in the South goes hand in hand with turning points in the narrator's life where tragic events involving family members shape one's life path. A similar picture emerges when one looks at education and learning where events forcing the narrator to drop out of school or finish some basic education are experienced in the South at the same rate. Moreover, negative financial and health shocks occur in the South at a much higher rate than elsewhere, painting a more nuanced picture of the rich life histories of those living in the US South in the late 19th and early 20th century.

A.4 Comparing Model Accuracy - Human Audit of Life Satisfaction Scores

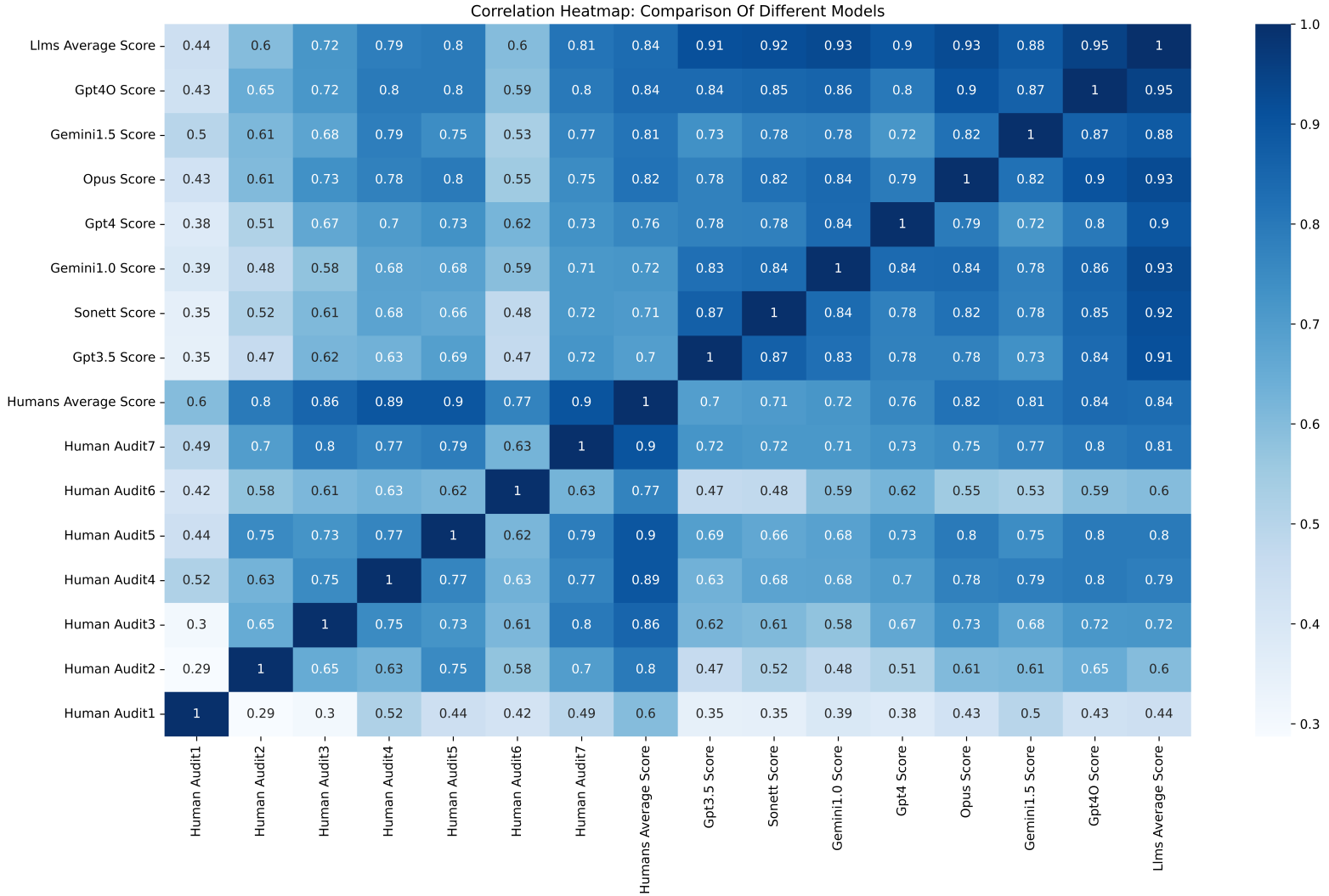
In the main text, we showed that the best LLM (GPT-4o) outperforms most of our human research assistants in terms of agreement with the modal human score. Here we, examine further the similarity and differences in scores across different models and evaluators. Figure A1 shows the correlation coefficients. Human evaluators have a range of coefficients from 0.29 to 0.79; the LLMs agree with each other with correlation coefficients 0.47 - 0.9. When we compare humans with the modal human score (including the evaluator themselves), the range of correlations is 0.6-0.9; for the LLM compared with the human average score, it is 0.7-0.84. Interestingly, the human auditor (No. 1) with the lowest agreement scores compared to other humans also ahs the lowest correlations with the LLMs, while the human auditor with the highest correlations with other humans (No. 7) has the highest correlations with the LLMs. This suggests that LLMs are particularly good at replicating the average or most typical human response. Our results also show that the heterogeneity of evaluations among the research assistants is in the same broad range as

gan, Indiana, Illinois, Wisconsin, Missouri, Iowa, Minnesota, North Dakota, South Dakota, Nebraska, and Kansas. The South is Delaware, Maryland, District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida, Kentucky, Tennessee, Mississippi, Alabama, Arkansas, and Louisiana. The West comprises Washington, Oregon, California, Nevada, Idaho, Utah, Montana, Wyoming, and Colorado. Finally, the Southwest is Arizona, New Mexico, Oklahoma, and Texas.

that of the LLMs; the worst human evaluator deviates more from other human evaluators than the "worst" LLM. Remarkably, the best LLM (GPT-4o) shows the same agreement with the human average as the LLM average, meaning that any "ensemble method" of averaging LLM scores is not outperforming the best frontier models.

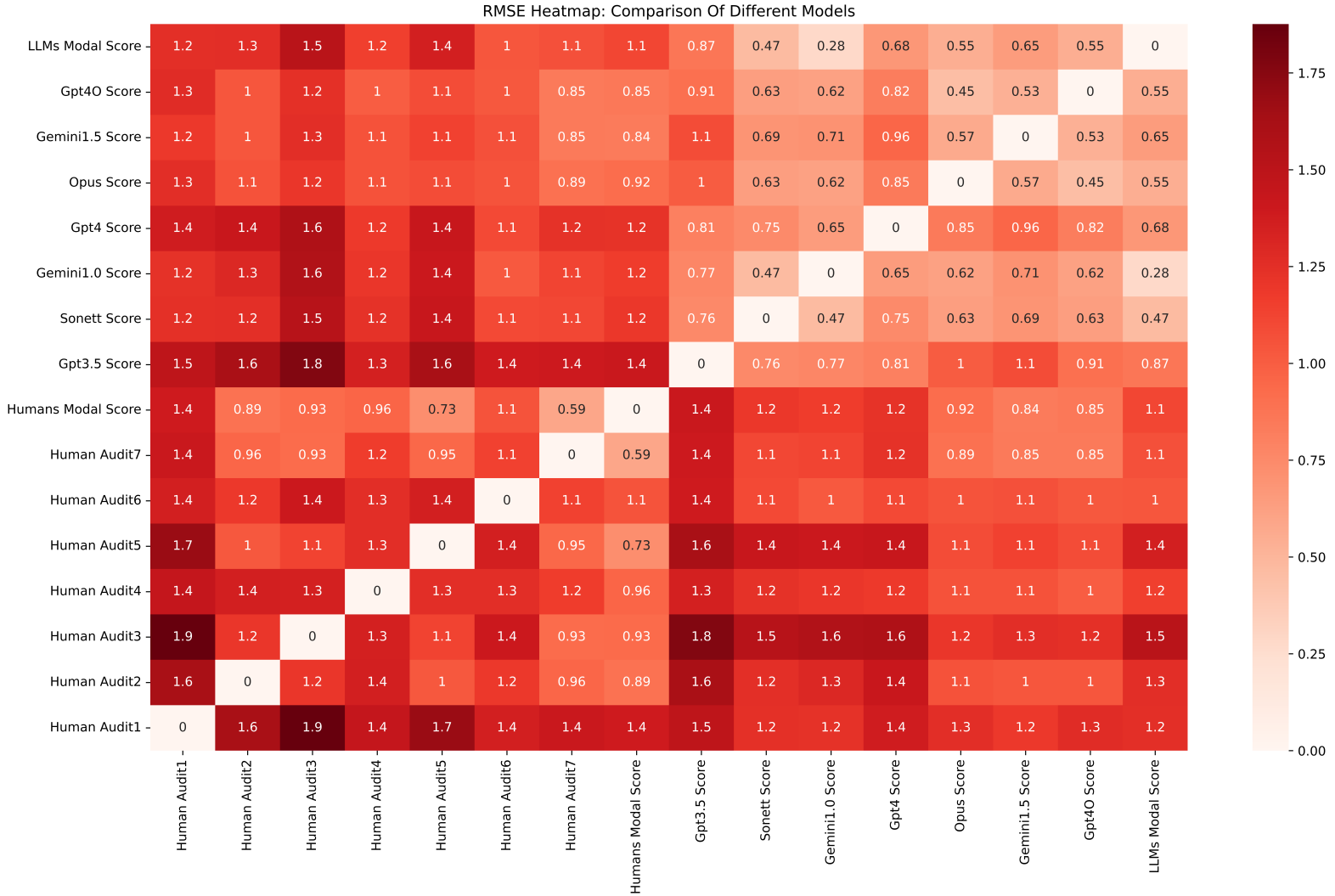
Figure A2 repeats the exercise using the root mean squared error as an accuracy indicator. Results are consistent with the pattern we found for correlations, with GPT-4o again scoring near the very top of LLM performance; only Gemini 1.5 shows a marginally better result.

Figure A1: All Model Comparison - Correlations



Note: The figure shows a heatmap of correlation coefficients for different human auditors and LLM models. Note that the human (LLM) average contains scores from all models, so that comparisons of correlation coefficients between individual auditors (LLMs) and the human (LLM) average are upward biased.

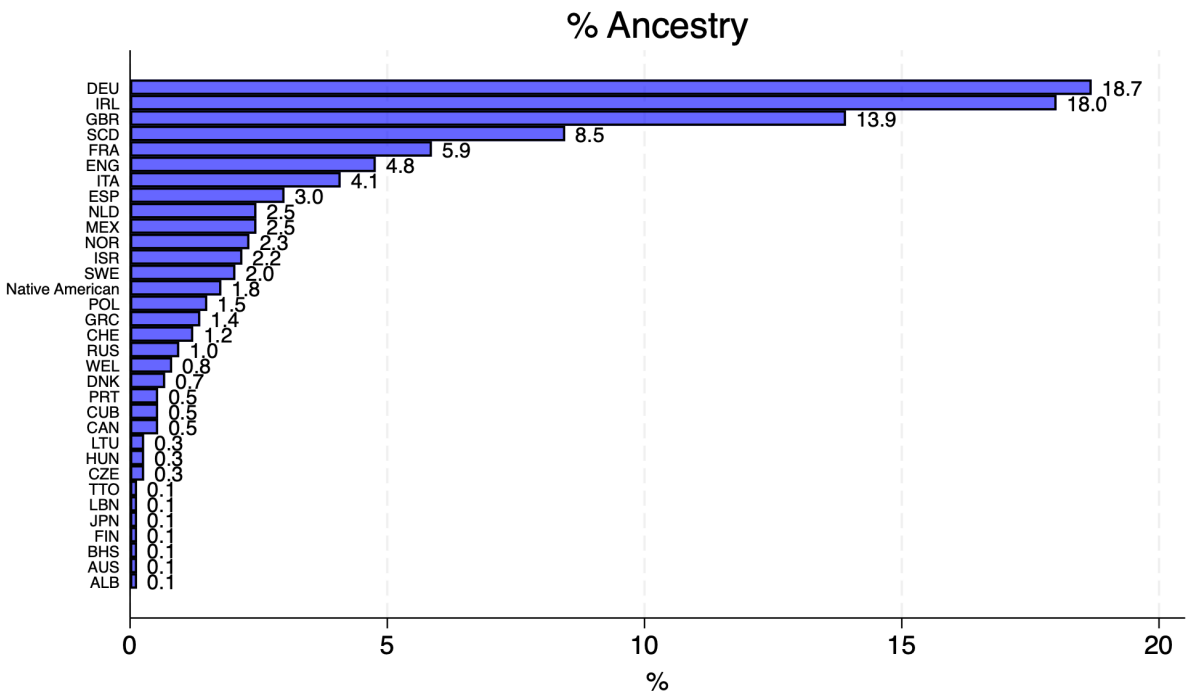
Figure A2: All Model Comparison - RMSE



Note: The figure shows a heatmap of root mean squared errors (RMSE) for different human auditors and LLM models. Note that the human (LLM) average contains scores from all models, so that comparisons of correlation coefficients between individual auditors (LLMs) and the human (LLM) average are upward biased.

A.5 Appendix Tables and Figures

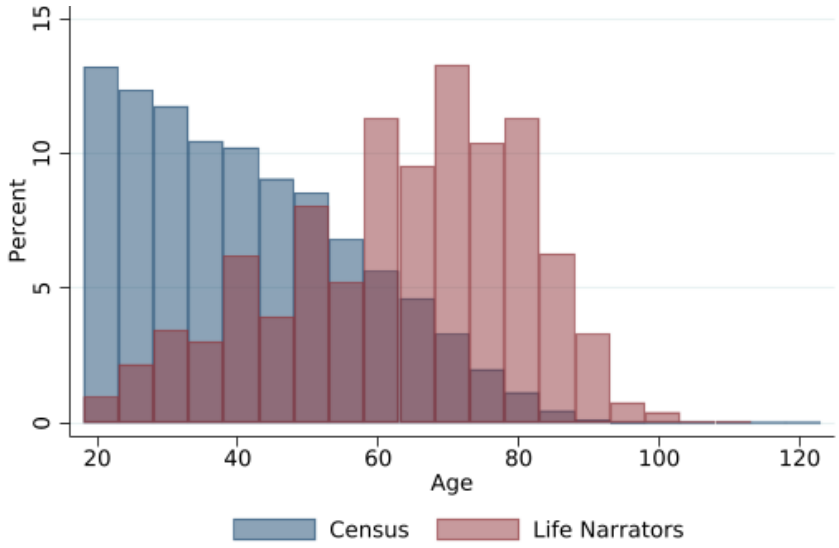
Figure A3: Ancestry of Life Narrators



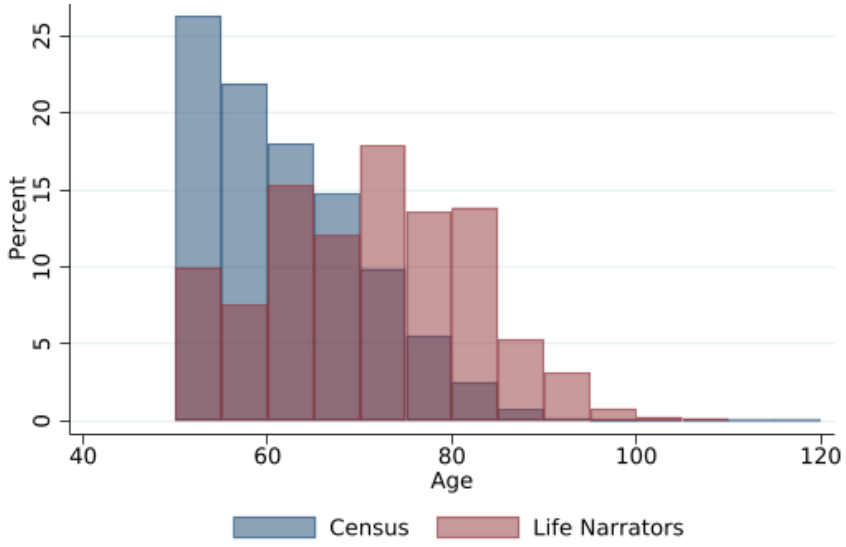
Note: This bar chart plots the ancestry of Life Narrators obtained by GPT and human coders. Those with American ancestry (irrespective of race) are excluded.

Figure A4: Age Distribution

(a) All Adults

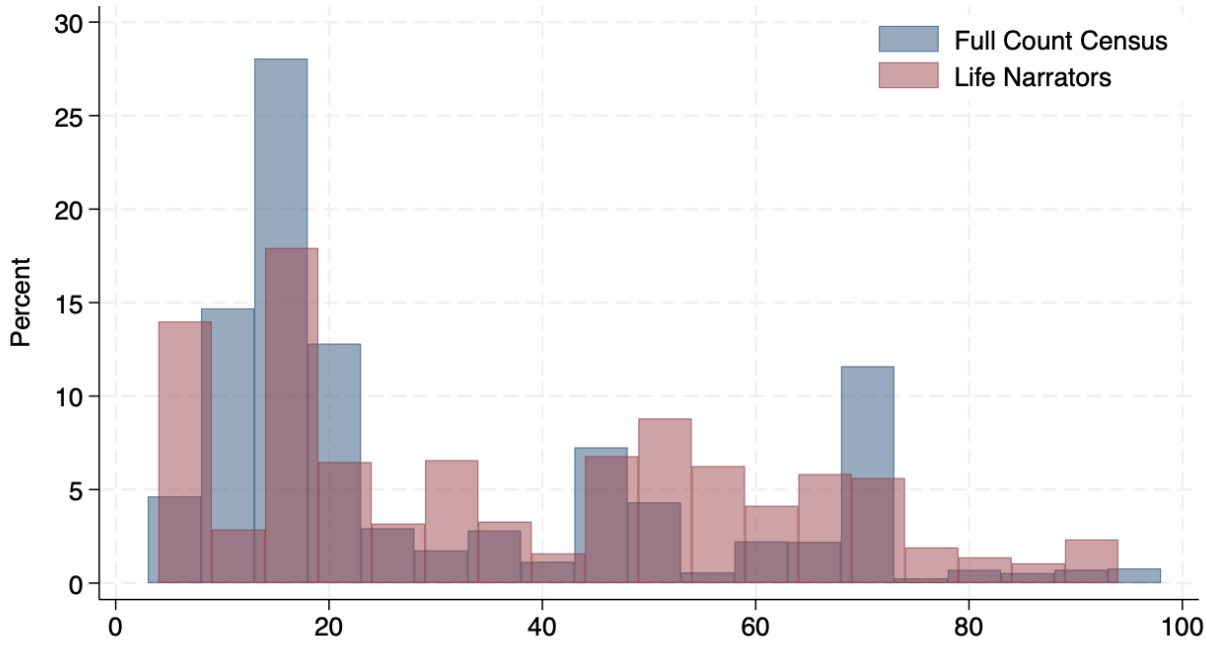


(b) Household Heads Over 50



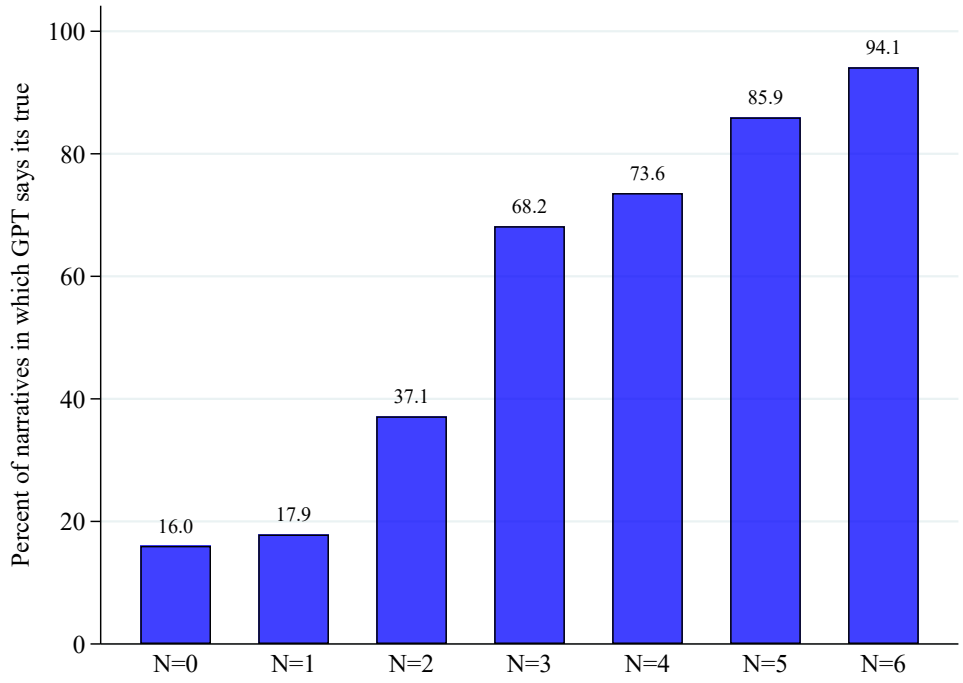
Note: This figure compares the age distribution of Life History Narrators with the the 1940's Census (Ruggles et al., 2024). Panel A4a plots all individuals older than 18 years old, while Panel A4b compares Life Narrators above 50 years old with Census household heads above 50 years old. Bin width was set to 5 years.

Figure A5: Distribution of Duncan Socioeconomic Index - Household Heads Over 50



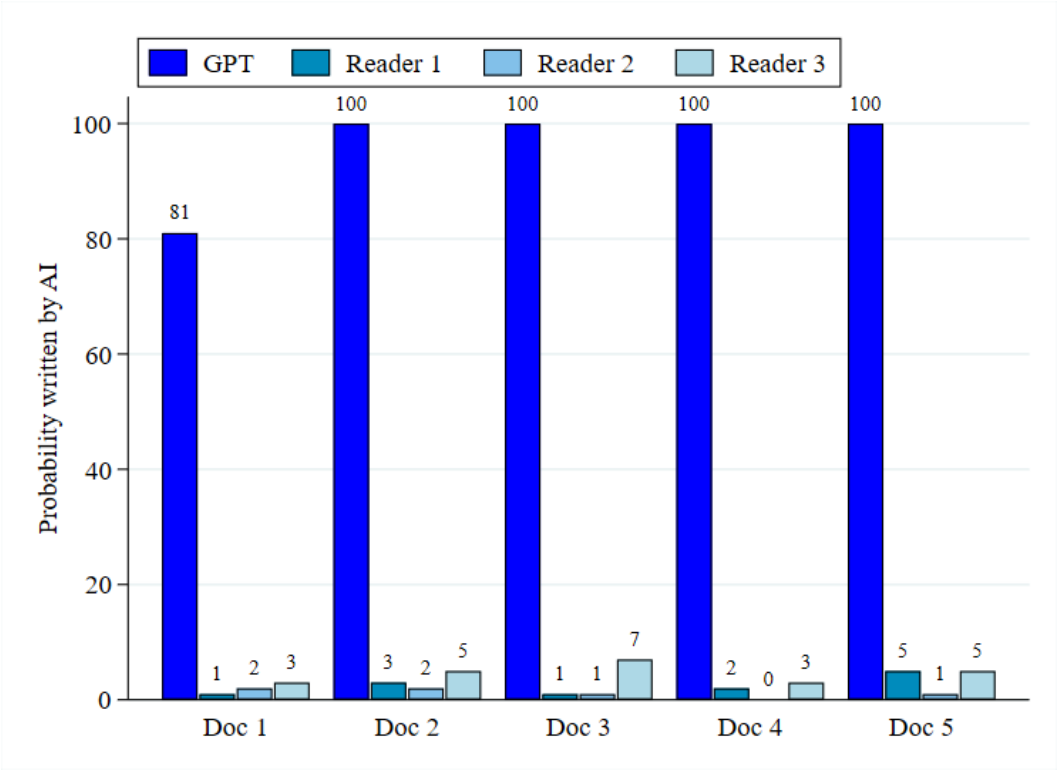
Note: This figure compares the Duncan Socioeconomic Index distribution of Life History Narrators with the 1940's Census (Ruggles et al., 2024). Panel A4b compares Life Narrators above 50 years old with Census household heads above 50 years old. The value for each Life Narrator was determined to align most closely with their main occupation during their prime years. Bin width was set to 5 years. Missing values, coded in (Ruggles et al., 2024) as SEI equal to 0 are not shown.

Figure A6: Conditional Probability of GPT Finding Hard Work to be Important



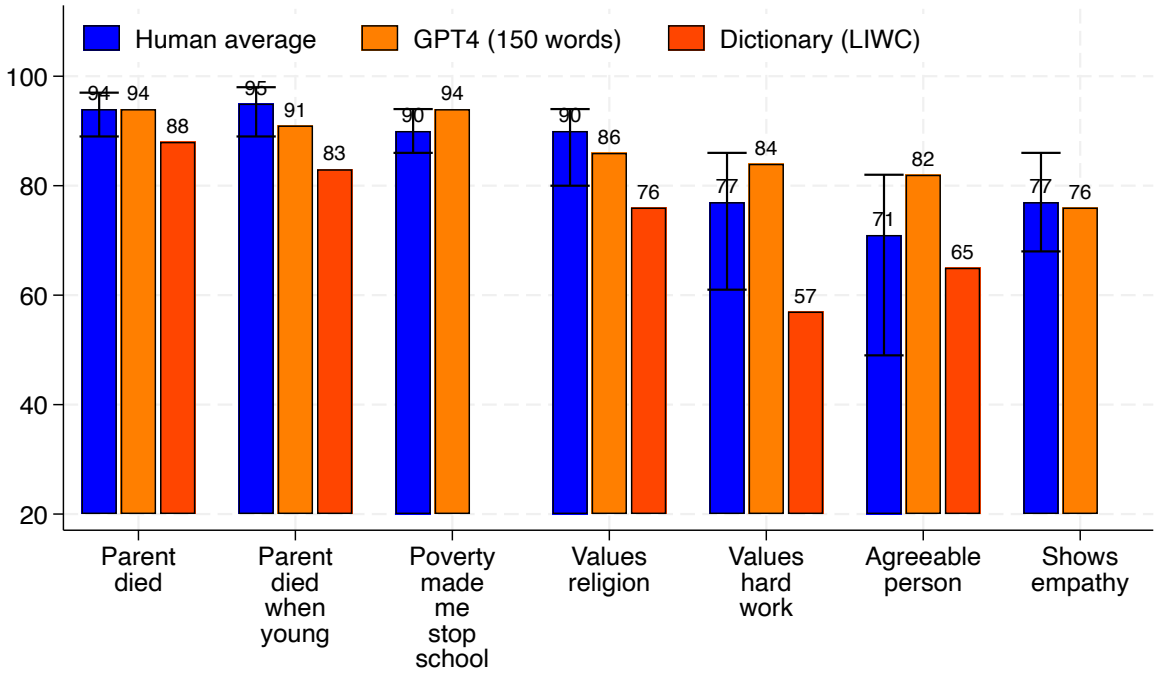
Note: This figure plots the probability that GPT finds hard work to be important conditional on the the number of human readers (N) finding hard work to be important.

Figure A7: Probability That a LLM Wrote Answers



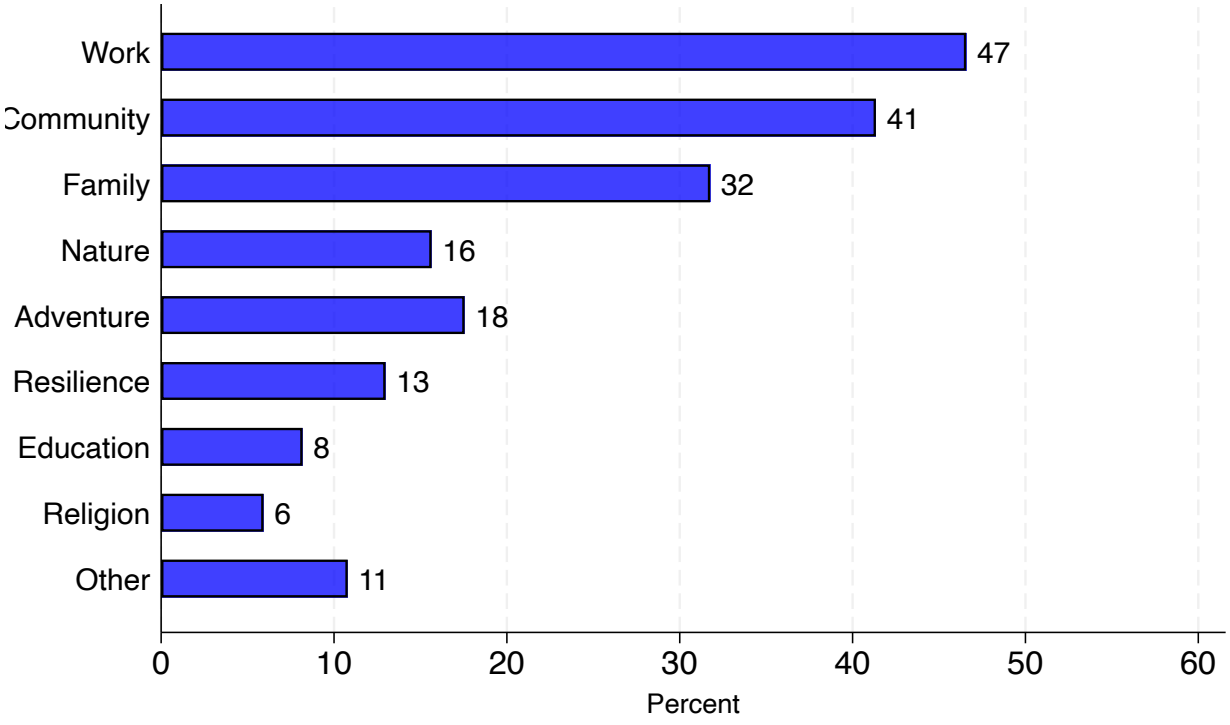
Note: This figure plots the probability that answers to the question about whether the narrator's life had happiness, satisfaction, or meaning were written by a LLM, according to GPTZero. The answers are shown for five randomly selected documents and the text answers provided by GPT4 and the three human readers.

Figure A8: Turing Tests: Computer vs Human Readings, Alternative Approach



Note: This figures compares the accuracy readings of the 300 randomly selected narratives by 6 human readers, GPT4, and a dictionary method. The score for each characteristic and narrative is either 1 (e.g. parent died) 0 (not present). This approach reimagines GPT as an additional Research Assistant, integrating its input into the analysis by comparing its scores to a common modal value. So, the modal value is calculated based on the responses of six human readers alongside GPT's own responses. In essence, this represents an alternative, less demanding Turing test, where the focus shifts to comparing the LLM's performance against the collective mode derived from all participants, including the LLM itself.

Figure A9: Sources of Life Meaning: Weighted Frequencies



Note: Weights are constructed from the 1940's full count Census considering the share of individuals above age 50 in each state by gender, race, immigrant/native and urban status. Out of 774 possible combinations of the previous variables, 198 are found among narrators.

Figure A10: Sources of Life Meaning: Frequency by Region

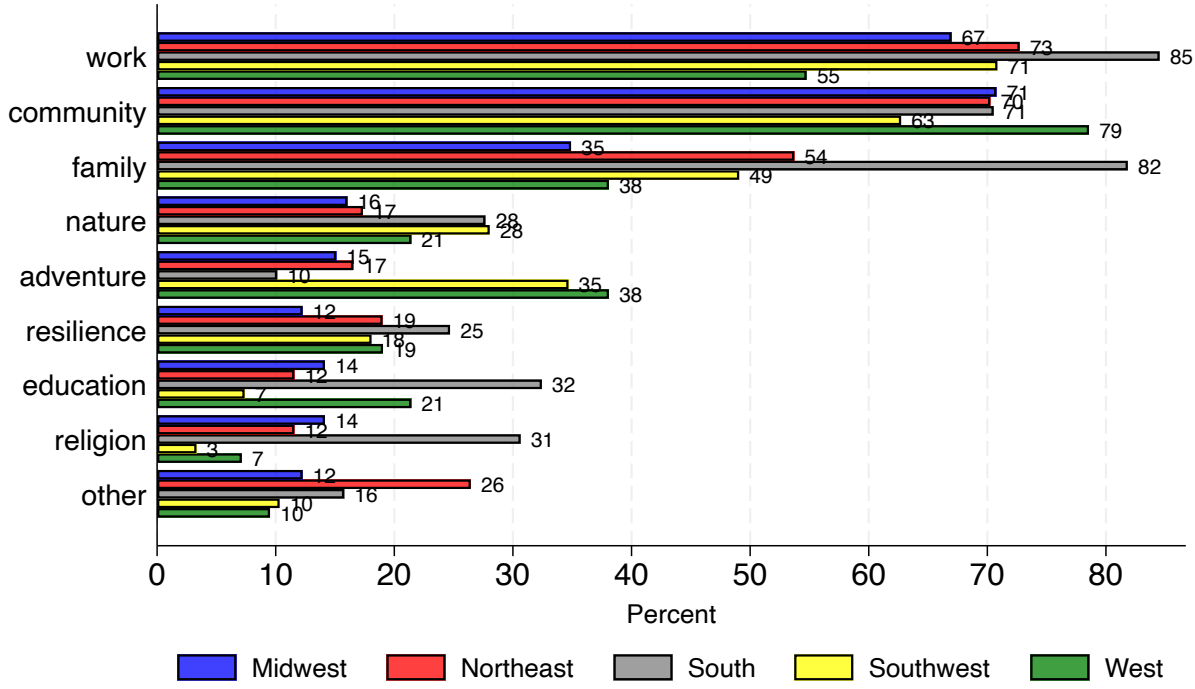
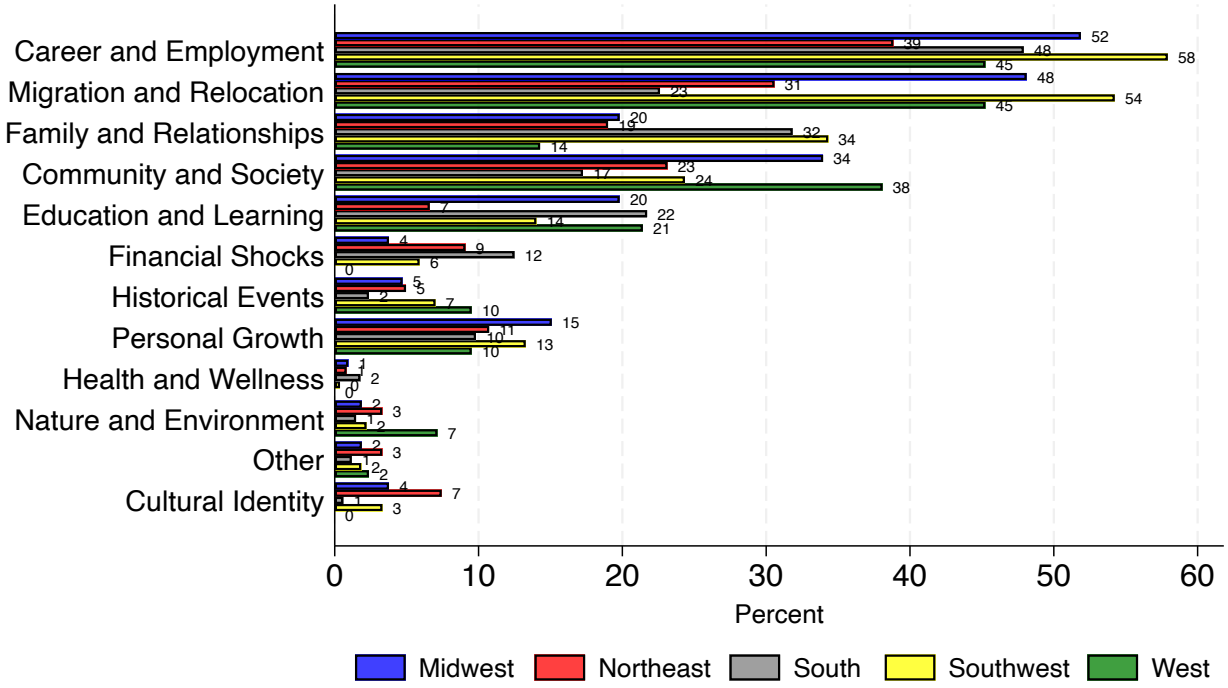


Figure A11: Critical Junctures: Frequency by Region

(a) Positive



(b) Negative

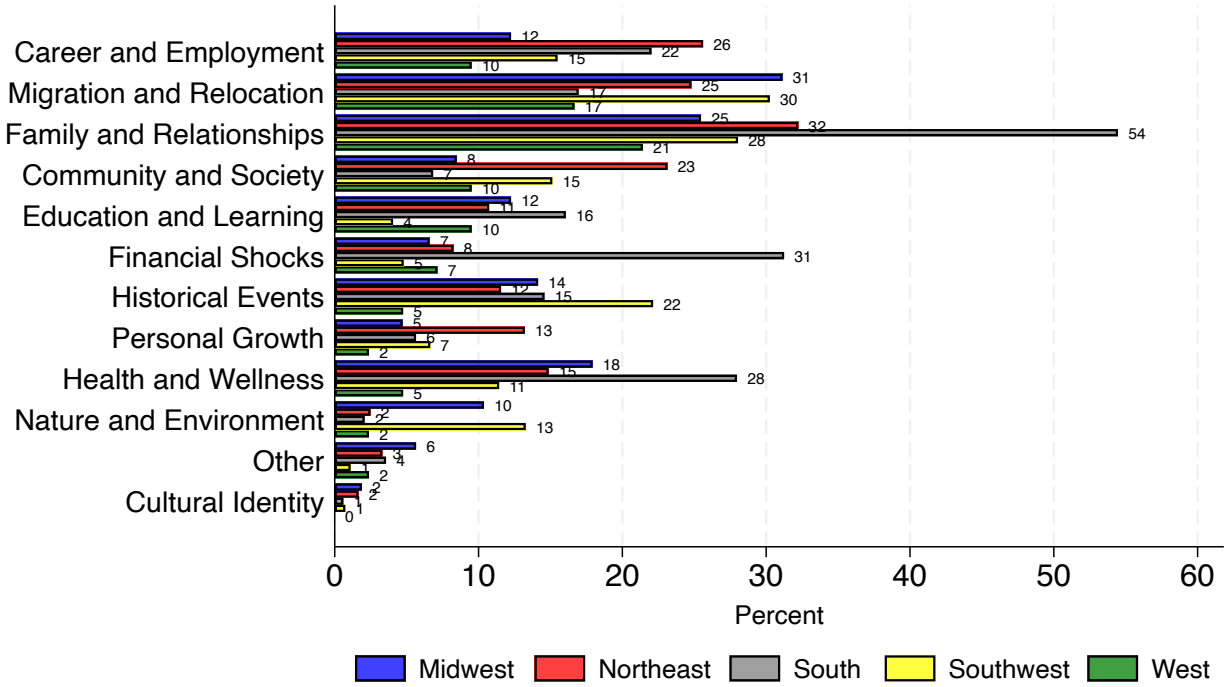


Figure A12: Sources of Life Meaning: Frequency by Subgroup

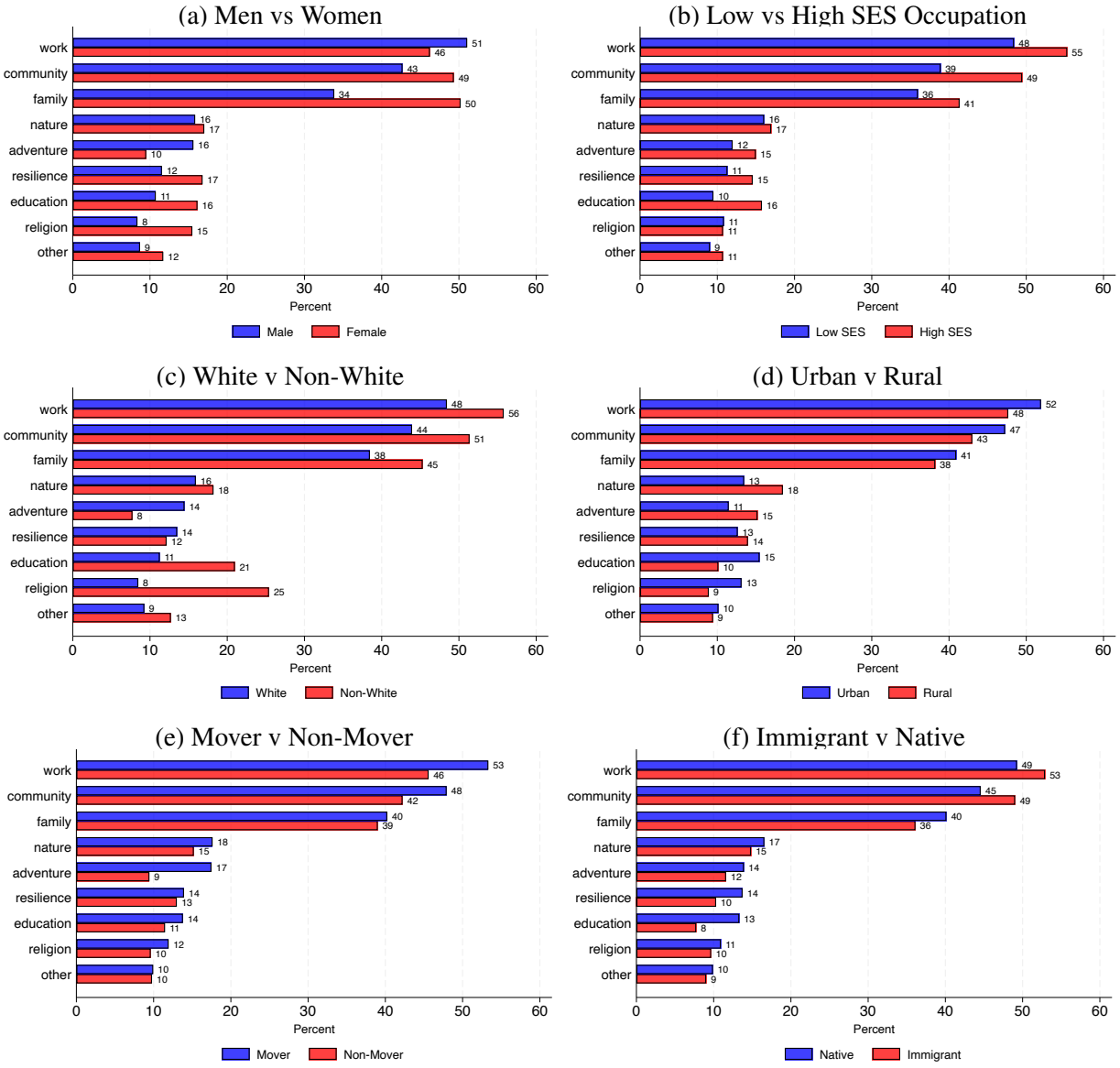
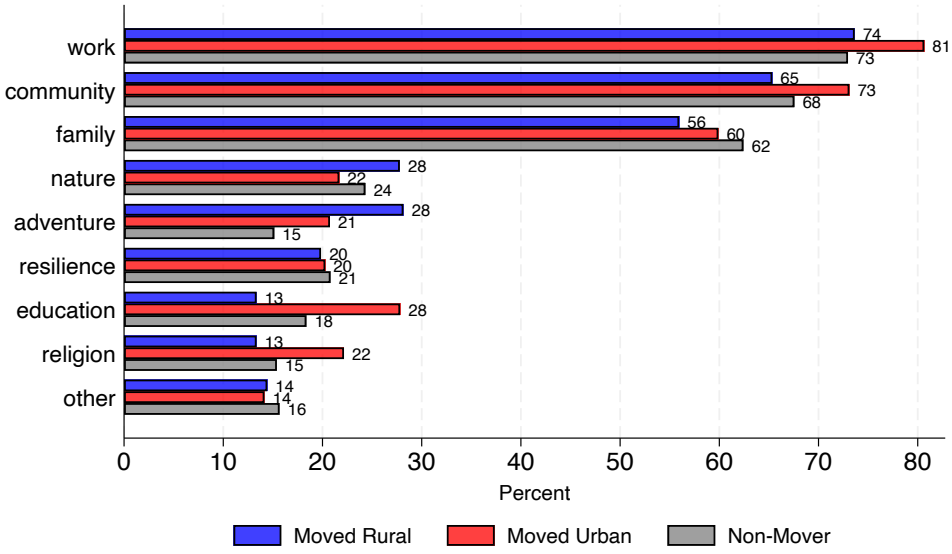


Figure A13: Source of Life Meaning: Frequency for Movers

(a) Moved Urban v Moved Rural



(b) Moved West v Moved Non-West

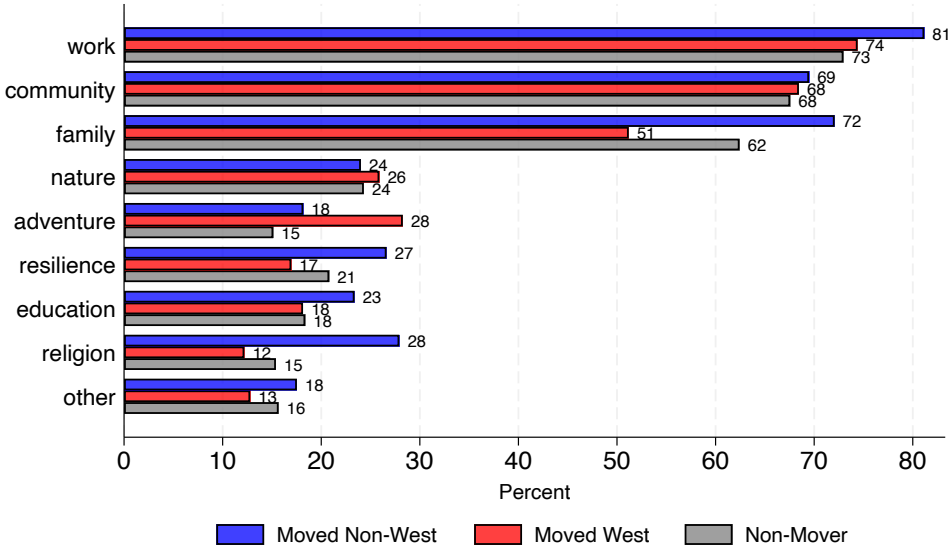


Figure A14: Sources of Life Meaning by Race and Migration Status

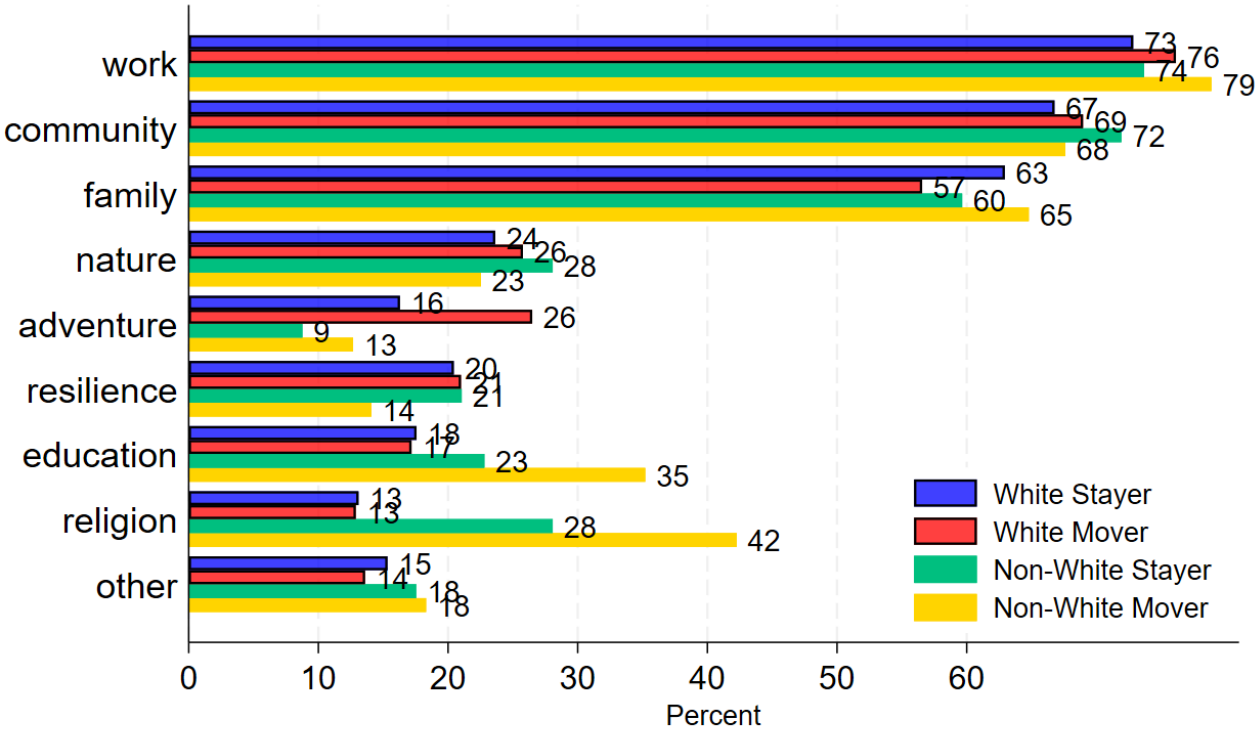
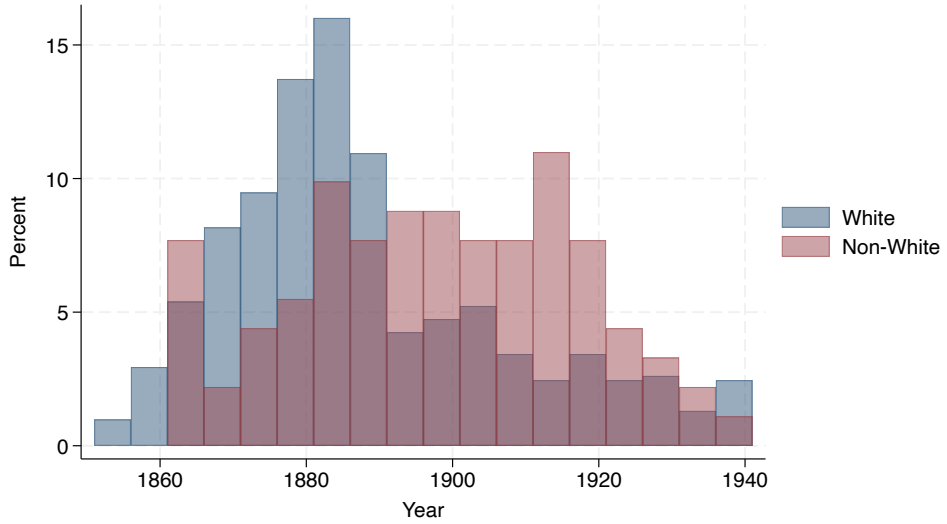


Figure A15: Movers by Race

(a) Year of First Move



(b) Change in Manufacturing Share

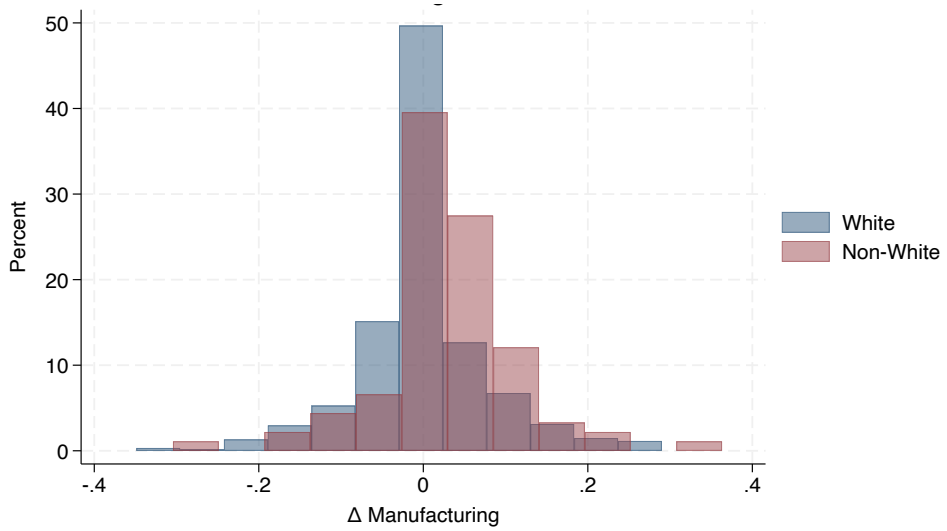
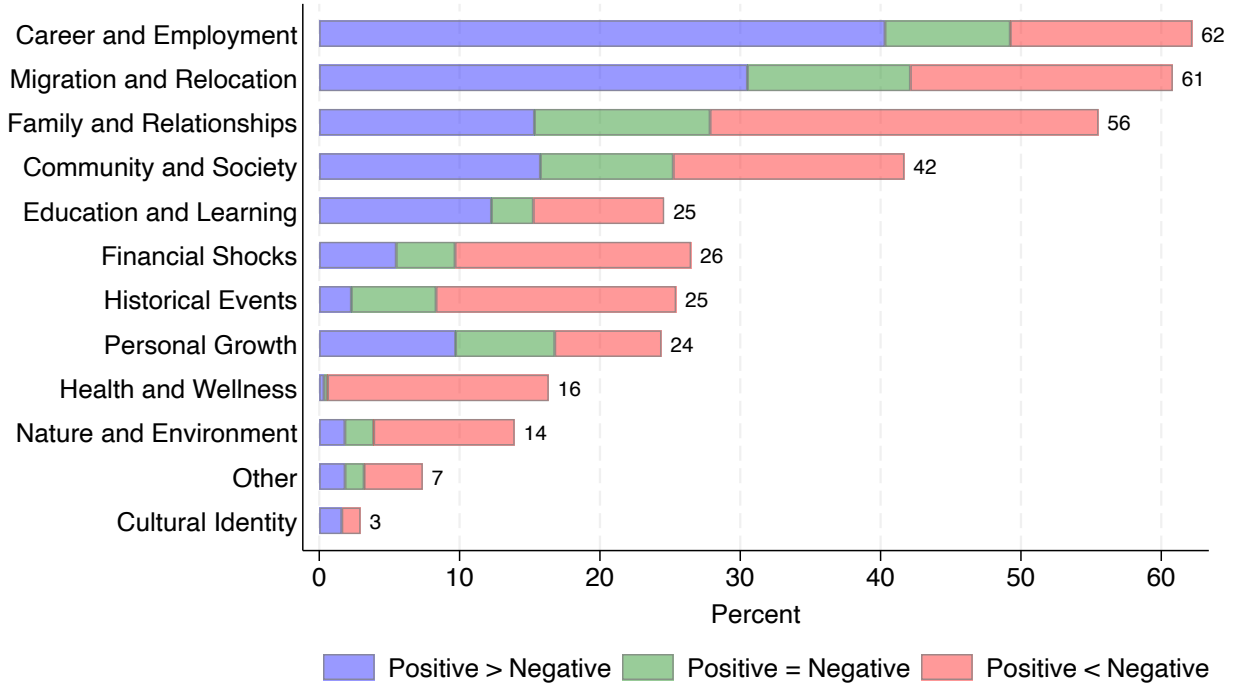


Figure A16: Critical Junctures: Weighted Frequencies



Note: Weights are constructed from the 1940's full count Census considering the share of individuals above age 50 in each state by gender, race, immigrant/native and urban status. Out of 774 possible combinations of the previous variables, 198 are found among narrators.

Figure A17: Clustering Dendrogram

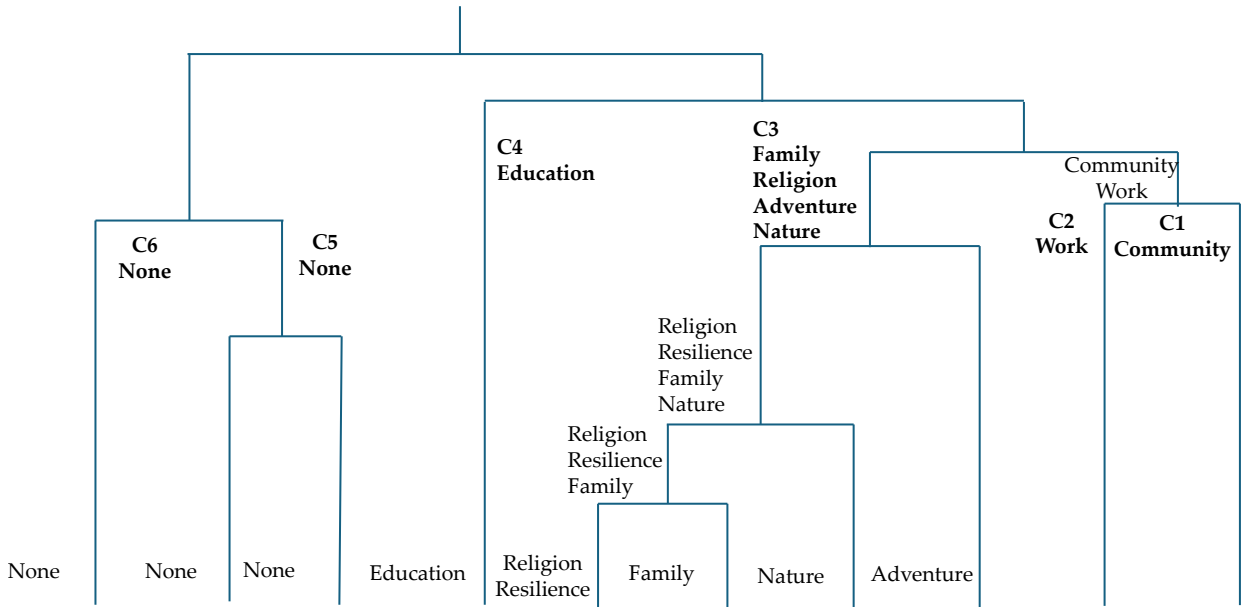


Table A.1: Turing Tests: Computer vs Human Readings, Alternative Approach

Percent of Narratives with Modal Human Score		
	Human Average [min, max]	GPT 4o (500 words)
Happiness, Satisfaction, Meaning, or Purpose	75 [69, 79]	85
Family	73 [70, 94]	85
Work and Career	71 [64, 75]	87
Religious and Spiritual Beliefs	86 [82, 90]	97
Education and Learning	91 [86, 94]	91

Note: This approach differs from the table 6 in that, here, individuals' values are compared solely with the modal values of three other people, excluding themselves. Simultaneously, the GPT response is evaluated against the modal response of all other human readers. Since GPT is compared with an even number of readers, in cases of ties (e.g., a 2 vs. 2 split), GPT is always considered correct.

Table A.2: Transitivity in narratives

File1	File2	File3	File1 vs. File2	File2 vs. File3	File1 vs. File3	Transitive Prop.	Score1	Score2	Score3
SarahWall	Obeah	FrankSower	File2	File3	File3	Yes	2	4	6
AnnieHight	MrsEllaCox	FrankNicho	File1	File2	File1	Yes	5	5	3
MikePellet	MaryAllen	ThatWasaMa	File1	File2	File1	Yes	4	3	3
MrsEllaDav	DennisPoti	GenevaStre	File2	File3	File3	Yes	5	5	5
GeorgeandB	GeorgeLins	MrsWTBoone	File1	File2	File1	Yes	5	4	4
MrandMrsWM	HardyJones	JHJakeByle	File2	File3	File3	Yes	4	5	4
JJJackson	OldSettler	SamJamesWa	File2	File2	File3	N.D	2	5	5
TakingCare	JimTurpin	GirlhoodLi	File1	File3	File3	Yes	4	4	5
JohnHRober	JimMiller	MrsJuanVal	File2	File3	File3	Yes	4	6	6
AFrenchSto	IrishCookB	ExStonecut	File2	File3	File3	Yes	3	5	5
WHThomas	Recollecti	MrsMikeMar	File1	File3	File1	N.D	6	3	4
SadieOrcha	TheHowes	AChangeofV	File1	File3	File1	N.D	5	2	5
LillieCrai	JDMashburn	WidowBucki	File2	File2	File3	N.D	2	5	3
MarySmith	JohnZMeans	JohnWHartm	File2	File2	File3	N.D	2	6	3
OTCardwell	Memoriesof	Thestoryof	File1	File2	File1	Yes	5	5	4
JackWPatte	Adventures	AmericanMe	File1	File3	File3	Yes	4	3	4
JamesCape	EESteen	Intervieww	File2	File3	File3	Yes	3	5	5
TheMarshal	FanniePerr	LouisFabri	File1	File3	File3	Yes	5	5	6
MeteorHell	RobertWill	MrEdMcCull	File2	File2	File3	N.D	5	5	5
SallysPrem	DrWhitelaw	MrsHCGates	File1	File3	File3	Yes	3	4	5

Table A.3: Odds Ratios for Sources of Meaning and Critical Junctures by Cluster

Panel (a): Sources of Life Meaning

Cluster	1	2	3	4	5	6
Family	0.1	1.1	4.2*	1.8	0.0	0.1
Work	1.3	2.0*	2.0	1.8	0.0	0.0
Community	2.5*	0.1	2.4	1.9	0.1	0.1
Adventure	1.8	1.0	3.2*	1.0	0.1	0.1
Nature	0.7	1.6	3.0*	1.5	0.0	0.1
Religion	0.4	0.8	2.2	4.4*	0.0	0.0
Resilience	1.4	0.7	2.1	2.3*	0.1	0.2
Education	0.0	0.0	0.0	55.3*	0.0	0.0

Panel (b): Positive Critical Junctures

Cluster	1	2	3	4	5	6
Migration	1.4*	1.1	1.0	1.0	1.2	0.5
Family	1.0	1.1	1.2*	0.8	0.8	1.1
Career	1.2	1.3*	1.0	1.2	1.2	0.3
Education	0.6	0.9	0.7	2.7*	0.9	0.9
Historical	1.4	0.9	0.6	0.6	2.1*	0.6
Community	1.5*	0.6	1.1	0.8	1.0	0.7
Financial	1.2	1.2	1.1	0.7	0.6	1.3*
Health	1.6	4.2*	0.1	2.5	0.3	0.8
Personal	1.4	1.7*	0.9	1.2	0.7	0.6
Cultural	0.8	0.5	1.3	0.7	0.5	2.4*
Nature	2.1*	1.5	1.2	0.7	0.6	0.5

Panel (c): Negative Critical Junctures

Cluster	1	2	3	4	5	6
Migration	0.6	1.1	1.1	0.9	0.6	1.9*
Family	0.7	0.9	1.0	1.1	0.6	2.0*
Career	0.7	1.0	0.9	1.2	0.7	2.0*
Education	1.0	0.7	1.2*	0.9	1.2	0.6
Historical	0.8	0.9	1.1	1.0	1.3*	0.7
Community	1.2	0.9	0.9	0.7	1.7*	0.5
Financial	0.8	0.8	0.8	0.9	1.2	1.5*
Health	0.8	1.0	1.0	0.8	1.0	1.3*
Personal	0.9	1.0	1.9*	0.9	0.9	1.4
Cultural	0.5	0.9	0.7	1.0	0.8	2.7*
Nature	1.2	0.3	0.8	0.9	1.9*	0.7

Note: This table reports the odds ratios for each source of happiness and critical juncture, positive and negative, for each cluster. A * indicates the highest odds ratio for each source of meaning or critical juncture.

Figure A18: Odds Ratio for Characteristics by Group without Interviewer FEs

Source	Female	Non-White	High-SES	Immigrant	Urban	Mover
Parent Died	1.63***	1.28	0.96	0.70*	1.01	1.16
Parent Died Young	1.25	1.15	0.83	0.79	1.11	1.12
Poverty Ended School	3.40***	1.14	0.37***	0.59	1.29	0.32***
Hard Work Important	0.72**	1.30	0.57***	1.46*	0.98	1.09
Religion Important	2.66***	2.68***	1.19	1.11	1.39**	0.71**
Agreeable Personality	2.00***	1.56**	1.31*	0.96	1.41**	0.69**
Showed Empathy	2.54***	1.65**	1.11	0.91	1.37**	0.64***

< 0.5 | $0.5 - 0.8$ | $0.8 - 1.2$ | $1.2 - 2$ | > 2

Note: Each cell in Figure A18 reports the odds ratio for each characteristic and subgroup, defined as the probability that narrators within each group mention a characteristic relative to the probability those *outside* the group mention it. The groups are women (relative to men), non-whites (relative to whites), high-SES (Duncan's SEI above 20, relative to 20 or below), immigrants (relative to native-born), urban (relative to rural), and movers (relative to stayers). The odds ratios, derived from logistic regressions, control for the log number of words. Stars represent the range of p-values of the tests that the likelihoods are the same based on robust standard errors, where *** means 1%, ** means 5%, and * means 10%.

Figure A19: Odds Ratios by Group without Interviewer FEs

Source	Female	Non-White	High-SES	Immigrant	Urban	Mover
Work	0.81*	1.39*	1.18	1.29	1.03	1.30**
Community	1.35**	1.39*	1.40**	1.34	1.03	1.19
Family	2.14***	1.36*	1.12	0.91	0.96	0.98
Nature	1.11	1.17	0.95	0.94	0.58***	1.13
Adventure	0.56***	0.48**	1.19	0.86	0.63**	1.97***
Resilience	1.60***	0.87	1.21	0.77	0.77	1.02
Education	1.65***	2.13***	1.64**	0.57*	1.45**	1.18
Religion	2.09***	3.85***	0.89	0.93	1.38*	1.22

Note: Each cell in Figure A19 reports the odds ratio for each source and sub-group, defined as the odds that narrators within each group derive meaning from a given source, compared to the odds of those *outside* the group finding meaning in it. The groups are women (relative to men), non-whites (relative to whites), high-SES (Duncan's SEI above 20, relative to 20 or below), immigrants (relative to native-born), urban (relative to rural), and movers (relative to stayers). The odds ratios, coming from logistic regression, control for the log number of words. Stars represent the range of p-values of the tests that the likelihoods are the same based on robust standard errors, where *** means 1%, ** means 5%, and * means 10%.

Figure A20: Positive Critical Junctures without Interviewer FEs

Source	Female	Non-White	High-SES	Immigrant	Urban	Mover
Career and Employment	0.28***	0.77	1.52***	1.00	1.19	1.95***
Migration and Relocation	0.73**	0.52***	1.18	2.21***	0.72***	3.35***
Family and Relationships	1.37**	0.74	0.94	0.92	0.71**	1.60***
Community and Society	0.90	1.26	1.46**	0.75	1.11	1.24
Education and Learning	0.96	1.63**	1.96***	0.60*	1.42**	1.76***
Financial Shocks	1.07	0.62	0.81	1.21	0.88	0.82
Historical Events	0.64	1.19	1.14	0.23**	0.97	1.02
Health and Wellness	0.84	1.60	4.74	3.90*	1.22	0.66
Personal Growth	0.47***	1.83**	0.81	0.29**	0.87	1.23
Nature and Environment	0.37**	1.55	1.05	0.81	0.66	1.20
Cultural Identity	1.77	3.64***	0.81	1.94	0.90	0.54

Note: Each cell in Figure A20 reports the odds ratio for each critical juncture and sub-group, defined as the probability that narrators within each group mentions at least one positive critical juncture, compared to the probability of those outside the group mentioning it. The odds ratios, coming from logistic regression, control for the log number of words. Stars represent the range of p-values of the tests that the likelihoods are the same based on robust standard errors, where *** means 1%, ** means 5%, and * means 10%

Figure A21: Negative Critical Junctures without Interviewer FEs

Source	Female	Non-White	High-SES	Immigrant	Urban	Mover
Career and Employment	0.66**	0.99	1.01	0.98	1.61***	0.87
Migration and Relocation	1.32**	0.60**	1.05	2.37***	0.78*	1.79***
Family and Relationships	2.59***	1.61***	0.67***	0.78	1.09	0.77**
Community and Society	0.78	0.73	1.24	1.06	0.90	0.58***
Education and Learning	1.23	1.08	0.79	0.55*	0.89	0.56***
Financial Shocks	1.37**	1.17	0.62***	0.64*	1.33*	0.60***
Historical Events	0.91	1.03	1.00	0.67	0.51***	1.40**
Health and Wellness	1.19	1.07	1.05	0.85	0.97	0.70**
Personal Growth	0.72	1.35	0.96	1.40	1.06	0.34***
Nature and Environment	0.75	0.18***	0.60**	0.61	0.28***	1.50*
Cultural Identity	1.21	1.06	0.52	1.88	1.03	0.39*

< 0.5 | $0.5 - 0.8$ | $0.8 - 1.2$ | $1.2 - 2$ | > 2

Note: Each cell in Figure A20 reports the odds ratio for each critical juncture and sub-group, defined as the probability that narrators within each group mentions at least one negative critical juncture, compared to the probability of those outside the group mentioning it. The odds ratios, coming from logistic regression, control for the log number of words. Stars represent the range of p-values of the tests that the likelihoods are the same based on robust standard errors, where *** means 1%, ** means 5%, and * means 10%

Figure A22: Meaning of Life by Critical Juncture Presence without Interviewer FEs

Meaning of Life

Positive Critical Junctures

	Work	Adventure	Family	Community	Learning	Resilience	Nature	Religion
Career	2.97***	1.80***	0.94	1.32**	1.32	0.97	0.90	0.56***
Migration	1.19	1.86***	0.91	1.24*	0.67**	0.75	1.23	0.84
Family	1.34**	0.80	2.22***	1.35**	1.07	0.87	1.26	1.14
Community	0.81	1.01	0.94	1.97***	0.80	0.84	0.74	1.63**
Education	1.41**	0.73	1.56**	1.69***	5.68***	0.79	1.08	2.09***
Financial Shocks	2.26***	1.13	1.75**	1.20	0.83	1.05	1.13	1.13
Historical	0.42***	2.10**	0.61	1.18	0.70	1.07	0.32**	0.12**
Health	1.02	0.70	1.02	1.25	3.17	0.71	2.34	3.83**
Personal Growth	2.15***	2.36***	0.93	1.59**	1.38	1.19	1.47	2.19***
Nature	0.61	1.94	0.70	1.17	0.22	0.68	1.55	0.26
Culture	0.54	0.70	1.34	1.73	0.77	1.00	1.65	0.28

Meaning of Life

Negative Critical Junctures

	Work	Adventure	Family	Community	Learning	Resilience	Nature	Religion
Career	1.09	0.92	0.92	0.66***	1.06	1.22	0.92	0.93
Migration	0.84	1.33	0.82	0.83	0.63**	0.75	1.41**	0.66*
Family	1.09	0.74*	1.80***	0.73**	1.61**	1.65***	1.11	1.64**
Community	0.42***	0.92	0.47***	0.67**	0.26***	0.63*	0.60**	0.40**
Education	1.23	0.63	1.54**	0.81	1.77**	1.05	1.11	1.60*
Financial Shocks	0.91	0.47***	1.32*	0.75*	1.36	1.13	0.68*	1.50*
Historical	0.58***	0.85	0.78	1.07	0.80	1.05	1.05	0.55**
Health	1.11	0.67*	1.35**	0.95	1.28	1.36	0.93	1.99***
Personal Growth	0.69	1.12	0.92	0.74	0.62	0.58	0.81	1.23
Nature	0.87	1.56	0.50***	0.89	0.36**	1.76**	0.98	0.54
Culture	0.53	0.33	0.49	0.38*	0.37	1.00	0.58	0.44

< 0.5 | $0.5 - 0.8$ | $0.8 - 1.2$ | $1.2 - 2$ | > 2

Note: Each cell in Figure A22 reports the odds ratios of a critical juncture with respect to a meaning of life category. Odds ratios are obtained via logistic regression controlling for log number of words. The first table uses as outcome variable an indicator taking value 1 if narrators report a positive critical juncture. Conversely, the second table focuses on a indicator taking value 1 if they report a negative one.