Cohort Succession Explains Most Change in Literary Culture

Ted Underwood,[a] Kevin Kiley,[b] Wenyi Shang,[a] Stephen Vaisey[c]

a) University of Illinois, Urbana-Champaign; b) University of Iowa; c) Duke University

Abstract: Many aspects of behavior are guided by dispositions that are relatively durable once formed. Political opinions and phonology, for instance, change largely through cohort succession. But evidence for cohort effects has been scarce in artistic and intellectual history; researchers in those fields more commonly explain change as an immediate response to recent innovations and events. We test these conflicting theories of change in a corpus of 10,830 works of fiction from 1880 to 1999 and find that slightly more than half (54.7 percent) of the variance explained by time is explained better by an author's year of birth than by a book's year of publication. Writing practices do change across an author's career. But the pace of change declines steeply with age. This finding suggests that existing histories of literary culture have a large blind spot: the early experiences that form cohorts are pivotal but leave few traces in the historical record.

Keywords: cultural change; literary history; age–period–cohort models; structural equation models; cultural analytics; cultural evolution

CULTURAL change is a central puzzle in the social sciences, and researchers have developed many ways to explain it. But theories of change can be loosely organized into two groups that predict different patterns of diffusion.

Theories in the first group posit that cultures change when individual members encounter an influence that fosters a new pattern of behavior. In the simplest version, new practices are directly "transmitted" from one person to another (Lucas et al. 2020). A more complex theory acknowledges that behaviors may be sustained by "cues embedded in the physical and social environment" rather than mental representations (DiMaggio 1997:267). An economic downturn, for instance, may transform behavior by changing the environment that conditions it, without transmitting specific ideas. But whatever causes are posited, models of this type predict that cultural changes will be visible as changes across individual lives.

A second group of theories hypothesize that people acquire stable patterns of behavior early in life—if not quite fixed beliefs, then "durable, transposable dispositions" (Bourdieu 1990:53). While this theory implies that fundamental change within an adult life span is rare, it leaves a great deal of room for change through cohort succession. Continuity across generations is thought to depend on processes of personal development that transform as well as transmit. In historical linguistics, for instance, it is clear that children drive change by imperfectly reconstructing the implicit structures that organized their parents' speech (Andersen 1973; Meisel, Elsig, and Rinke 2013). Much language change happens "at a parent's knee." Political opinions also appear to be fairly stable after a formative period in youth (Alwin and Krosnick 1991; Sears and Funk 1999). Following Vaisey and Lizardo (2016), we characterize these explanations as theories of "acquired dispositions."

Because human history is notoriously complex, both theories described above may have some validity. Instead of trying to decide between them, a useful experiment would aim to set bounds on their respective importance in a particular domain. Even a rough sense of their relative importance could reshape research because researchers' attention is not at present equally distributed between these two modes of change. Instead, different theories dominate in different disciplines.

Linguists, for instance, have firmly established the centrality of cohort succession in language change. But if we turn from glacial processes of grammatical drift to a volatile phenomenon like the writing of fiction, the role of acquired dispositions becomes less clear. Can complex practices like modernism and detective fiction really be described as dispositions acquired in youth? Literary historians more commonly explain them as immediate responses to new literary models or historical events. It is proverbial, for instance, that modernism was a response to modern warfare (James 2013; Sherry 2003). "World War I was a great engine of modernism, endorsing the chaos of shattered belief" (Ramazani and Stallworthy 2012:1904). Historians do occasionally gesture at generational succession, for instance by speculating that "a new generation of poets" reacted against their precursors (Ramazani and Stallworthy 2012:1889). But it is harder to find evidence for this theory. The historical record is much more likely to document interactions between adults than to preserve young readers' early experiences of acculturation (Stearns 2008).

In short, histories of intellectual and artistic expression currently rely on a different theory of change than research on language or political affiliation. This disciplinary divide has been difficult to bridge because artistic dispositions are not easily captured by the close-ended survey questions used to measure change in social science. However, new modes of analysis provide new leverage on the problem. Machine learning can measure differences between periods and cohorts even in unstructured data like the corpora of images and texts that document artistic expression.

The Case of Fiction: Hypotheses and Data

We take the history of English-language fiction as a convenient case study. Works of fiction are preserved in large numbers, and a determined researcher can connect most titles to a year of first publication and most authors to a year of birth. We use topic modeling to identify 200 latent variables that organize a large corpus of English-language fiction, and then ask whether those variables are distributed by cohort—as a theory of acquired dispositions would predict—or by publication year, as we would expect if writers' practices are immediately shaped by events that affect multiple generations at once. Since we anticipate that both theories have some validity, we also preregister a hypothesis about the specific categories of cultural practice more likely to be governed by early socialization or by immediate influence.

Our topic model uses 29,341 volumes of fiction, written in English and published between 1880 and 1999. Analysis focuses on a subset of 10,830 volumes where records provide strong confidence in the date of first publication and the author's

year of birth. For the process of corpus construction, see "Data Construction" in the appendix.

We could have represented the corpus simply by measuring word frequencies. But the themes and genres of literary culture are more volatile than vocabulary alone would imply. In one volume, the word "rifle" reveals the influence of Western genre fiction; in another, it signals mainly that the book was written in 1944, in a world overshadowed by war. Topic modeling captures this contextual dimension of meaning, since the same words can be assigned to different topics in different contexts (Blei, Ng, and Jordan 2003).

Topic models do ignore word order. So if a literary practice affected only syntax and structure, without producing any ripple effects on diction, it would be invisible in a topic model. At first glance, it may seem intuitively likely that many aspects of literature could have this purely structural character. For instance, the boundary between novels and short stories isn't defined by anything we would normally call a "topic." Like novels, short stories can be written about love affairs in Tsarist Russia or about outer space. The difference between a novel and a collection of short stories—one might intuitively assume—is found purely in the ordering of words that creates global coherence in one case and local coherence in the other. So we might expect the difference between these forms to be invisible in a topic model. Differences of literary quality, likewise, may seem mostly structural. We don't normally say that good writers are good because they use better words, but because they put them in a better order.

But while it sounds reasonable to posit that content and structure are separable, the premise has not fared well in practice. Researchers have not been able to locate important boundaries in the history of fiction that aren't legible in diction. Even the purely formal distinction between short stories and novels leaves clear traces in word choice: story collections and novels can be distinguished with 80.1 percent accuracy using only the topic proportions in our model. In the online supplement associated with this article we examine the topics that make this distinction possible. Here it will suffice to observe that novels don't turn into short stories by subtracting sentences at random. Specific aspects of life get elided in the shorter form—like references to the passage of time and sprawling family structures. Other aspects of life receive more emphasis. So short stories do in practice cluster in particular topics. Judgments about literary prestige are likewise easy to predict with topic proportions, which align more closely with human judgments of quality than neural document embeddings do, even though the document embeddings attend to word order (van Cranenburgh, van Dalen-Oskam, and van Zundert 2019:637). We therefore follow other researchers in taking topics as a reasonable proxy for the latent social, stylistic, and formal variables organizing text (Barron et al. 2018; Lambert et al. 2020; Nelson 2017).[1] Of course, no proxy is perfect. But topic models have performed well across a wide range of literary questions; extensive testing has not yet revealed an important aspect of fiction they are unable to represent.

For details of the topic modeling process, including the rationale for selecting 200 topics, see "Topic Model" in the appendix. Once a model was trained, two readers manually labeled topics and assigned them to 10 categories. "Genres" like love stories and detective fiction were distinguished, for instance, from "aspects of

physical description" like mountain scenery and from "events" like wars. When readers were uncertain, they assigned topics by default to an "uncategorized" group. Comparing categories independently assigned by the two readers gave us a Cohen's kappa of 0.59.

We analyzed this model in two interlocking ways.

- Regression models partitioned the overall variance in the corpus, dividing it between birth year (cohort) and publication year (period) variables. This provided a picture of change at the population level.

- Structural equation models characterized longitudinal change within each author's career, distinguishing topics where writers tend to change in durable ways from topics where durable change is outweighed by other variation.

Both of our analyses represented books as distributions across topics in the same 200-topic model. But they drew on slightly different subsets of the corpus. We limited the regression model to 5,572 books by authors of confirmed U.S. nationality because the changing composition of the corpus might otherwise produce spurious period effects. The structural equation model was not limited to U.S. authors. But since it investigates longitudinal change, it was necessarily limited to authors who have composed more than two books; this gave us a subset of 10,355 books to study.

Change at the Population Level

Differences in behavior can be associated with time in three ways: as consequences of a subject's age, characteristics of a generational cohort, or effects associated with a historical period (Ryder 1965). In our regression models, these three independent variables attempt to explain a dependent variable that is the prominence of a particular topic across 5,572 books (the proportion of words assigned to the topic).

Age strongly constrains some behavior (e.g., driving) but is not very useful in predicting literature. (Books written by 30-year-olds today do not resemble books written by 30-year-olds in 1880.) Most variation associated with time is explained more effectively by the year a book was published (its period) or by the author's year of birth (its cohort). Following Vaisey and Lizardo (2016), we define a parameter δ that is the variance explained by cohort, divided by all variance associated either with period or cohort:

$$\delta = \frac{V(C)}{V(C) + V(P)}. \tag{1}$$

In estimating this parameter, we confront the challenge that the system of age, cohort, and period is perfectly collinear. If we know the year an author was born and how old the author was when a particular book was published, we immediately know (by addition) the year the book was published.

A common solution to this problem is to constrain the precision of variables. For instance, instead of using the original, numeric year of birth (or publication) in a regression, we can divide both timelines into segments and give all authors (or books) in a segment the same categorical value. This coarser representation of

time breaks the collinearity between age, cohort, and period (Mason et al. 1973). However, the solution remains open to an objection of arbitrariness. Timelines constrained with different degrees of granularity may produce different results. To model this system convincingly, we need some "reason to think that the identifying constraint imposed is correct" (Bell and Jones 2013:164).

But how can researchers know what level of constraint is correct for a given data set? One way to get leverage on this question is to consider the spurious precision of a collinear model as a special case of the more general problem called "overfitting." Because the age–period–cohort system provides redundant information, it permits a wide variety of solutions that may not generalize to new samples. That is, they have excessive variance. On the other hand, if we discretize time too coarsely, our representation may not provide enough information to fit sudden shifts in topic prominence. It will produce models with excessive bias.

This is a classic problem in machine learning, and the classic way to solve it is to search for model specifications that empirically minimize the sum of bias and variance on out-of-sample evidence. To do this, we discretized period and cohort variables with a window that could vary from 4 to 24 years in width. Exploring all of those options for both variables produced a six-by-six grid of possible model specifications. We cross-validated all 36 models for each of 200 dependent variables (topics), choosing in each case the specification that maximized r^2 on out-of-sample authors. Relatively large windows of 12 to 24 years were often preferred.

Age was represented as a quadratic polynomial, since it has a very weak and mostly linear relation to literary history. In fact, age is weak enough to suggest an alternate solution to the collinearity problem: we could explicitly assume that age plays no role here and drop it from the regression entirely. This alternate approach produced an estimate of δ very close to the one we report below (within two percent). Relatively large windows were still often preferred, suggesting that collinearity is not the only kind of overfitting these models need to avoid. See the supplementary material associated with this article for fuller description of alternate models.

Preregistered Hypotheses

Before performing any analysis on the topic model, we preregistered the sample (Underwood et al. 2021). We also agreed in advance to calculate δ for the whole corpus by adding up the variance explained by period and cohort for all topics, weighted by the number of words assigned to each topic. Large topics, and topics where period and cohort collectively explain a lot of variance, would count for more in this average than small topics and topics where total r^2 is small. We agreed to exclude "accidents of transcription" and "author-dominated topics" from the overall total; for a fuller account of these categories, see "Topic Categories" in the appendix.

We also preregistered a hypothesis that three categories of topics ("events," "accidents of transcription," and "technological changes") would collectively have lower than average δ parameters. We expressed this as a weak hypothesis about effect size because n is single-digit for each category and only 19 for all three taken
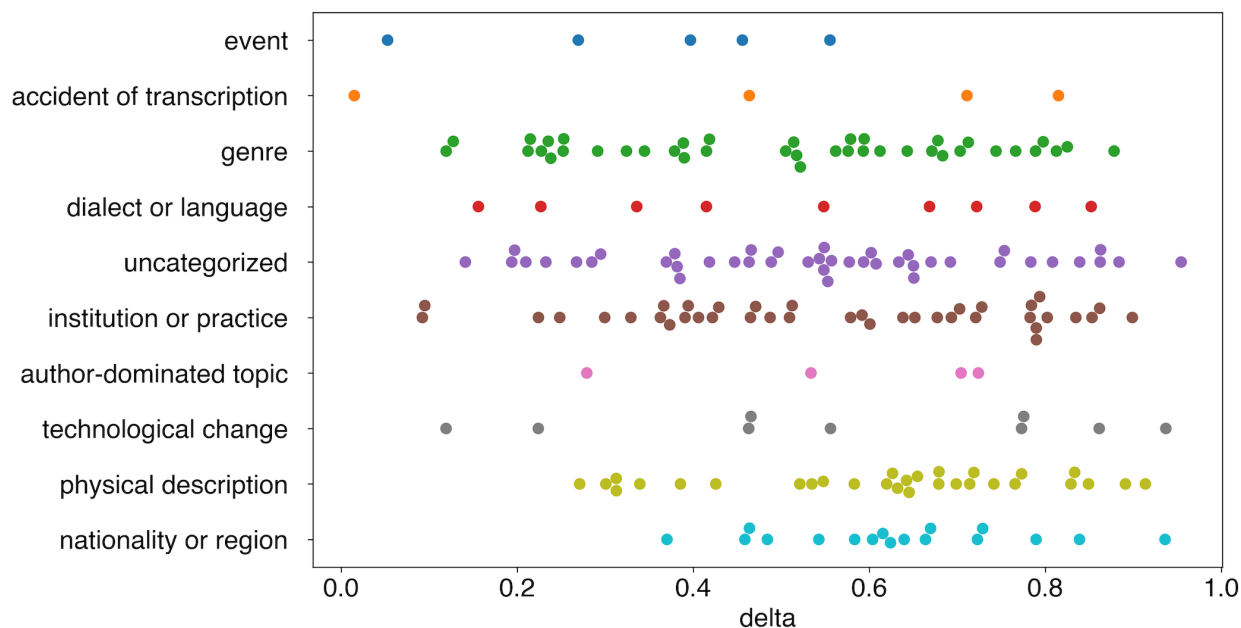
Figure 1: The delta parameters for different topic categories; total $n = 200$. When δ is greater than 0.5, topic prominence is more effectively explained by birth cohort than publication date. Categories are sorted by mean δ.

together. In this topic model, the events are mostly wars and revolutions; we expected these to be period-driven for obvious reasons. Accidents of transcription include optical character recognition (OCR) errors that are often shaped by printing practices in a particular decade. And mentions of a technology seemed more likely to correlate with the current prominence of the technology in society at large than with the author's experiences in youth.

Regression Results

The process of cross-validating to choose model specifications introduces variation, so we ran the regression model five times and averaged results. Adding up variance across topics and weighting for the size of each topic, total δ was 0.547. We use these averaged results in the analysis that follows. To estimate uncertainty, we also calculated δ for 50 bootstrap resamples of the 5,572 volumes being modeled; results ranged from 0.523 to 0.618, with a median of 0.547. In all of the models we ran, more than half of the variance associated with time was explained by an author's year of birth.

Delta parameters for individual topics ranged widely within categories (Figure 1). Our preregistered hypothesis that three categories would collectively have lower deltas was narrowly confirmed ($n = 19$, Cohen's $d = 0.28$, where we had hypothesized $d > 0.1$). But it was confirmed mostly because one "accident of transcription" and several "event" topics have notably low deltas. The top row of Figure 2 shows the typical chronological profile of a strongly period-driven "event"
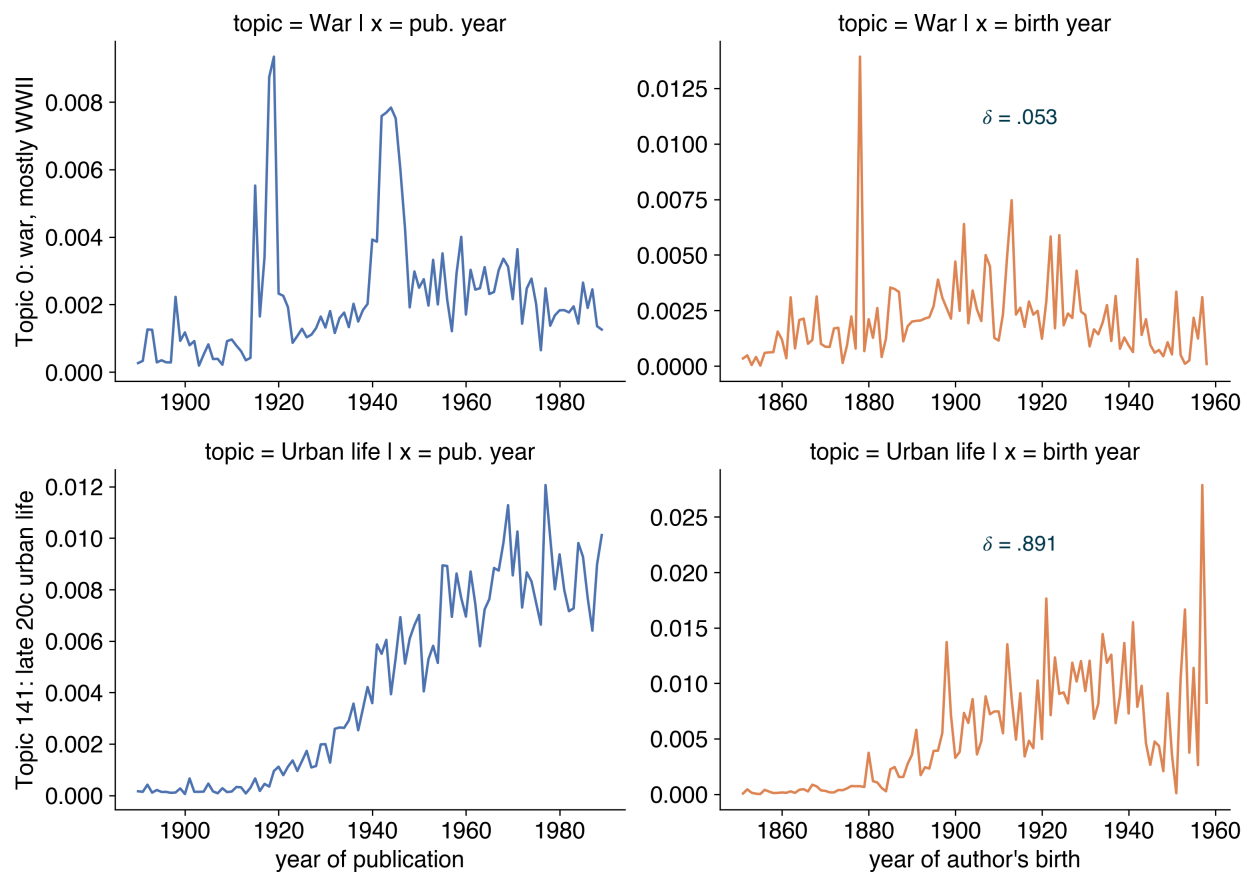
Figure 2: The yearly frequencies of two topics with markedly different delta values; $n = 29,351$. Delta will be low when changes unique to the left-hand column are both large and durable for out-of-sample authors; it will be high when the same is true on the right.

topic, associated with description of warfare and spiking for several years at both World War I and World War II. The predictive information in those broad spikes is not available when we plot the topic by birth year. (The sharp variations in the birth year plot are mostly due to single authors, and since model specification is defined by performance on out-of-sample authors, cross-validation learns to ignore them.)

"Technological changes," by contrast, were not as period-driven as we had expected. This surprise was largely caused by a division within the category. References to emergent technologies (like the automobile) were in fact period-driven. But we had also counted technologies in decline (like horse-drawn transportation) as examples of technological change, and these topics often associated more strongly with cohorts. This is an intriguing pattern, fairly clear in the technological category, but we cannot say that it holds generally across the model. If we loosely distinguish emergent and residual topics by measuring overall linear correlation with time across the century, the emergent topics (with a positive slope) do have lower deltas than the residual topics (with a negative slope). But the pattern is not strong ($r = -0.068, n = 200, p = 0.339$, two-tailed).
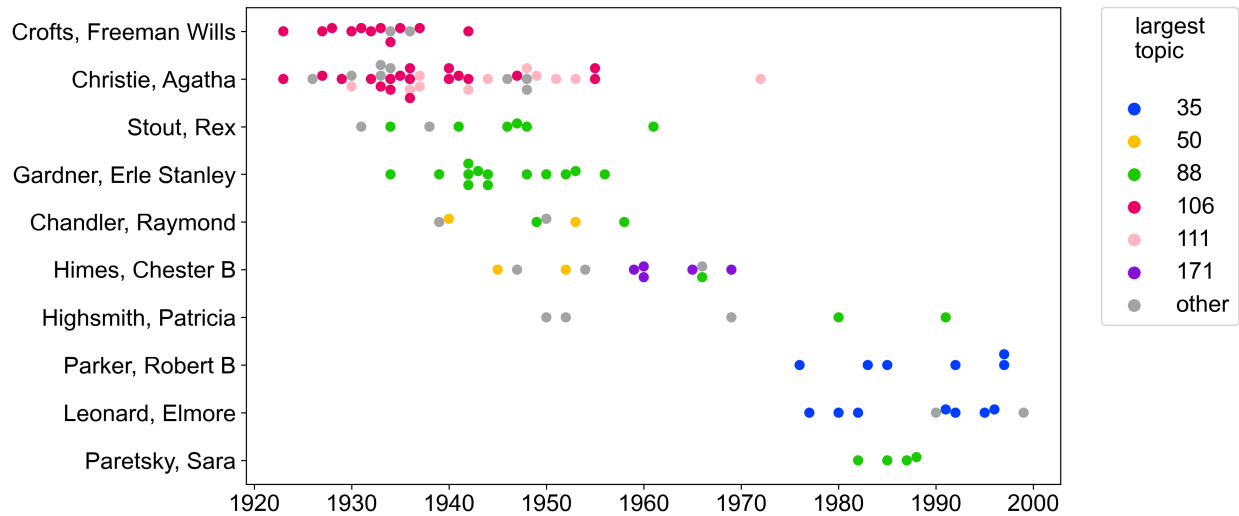
Figure 3: Books by 10 well-known authors of mystery and crime fiction, positioned by publication date and colored by their largest topic.

We mention this null result because it may be tempting to speculate that new literary practices are always introduced at first through events that transform a whole period, and become cohort-driven phenomena only in decline, because acquired dispositions persist in older writers. This is a plausible hypothesis, worth further study, but we have not yet found evidence that consistently supports it. In fact, it is easy to find dramatic counterexamples. In the bottom row of Figure 2, a topic describing late twentieth-century urban life (with words like *room*, *street*, *hotel*, and *bus*) is clearly emergent across the period modeled. But it has a delta of 0.891 because birth year is more informative here than publication date: a whole generation born in the late 1940s was distinctly less fond of this topic than their immediate precursors or successors.

To get an intuitive sense of the way cohorts define literary culture, it may help to glance at Figure 3, which visualizes all the books in our corpus by 10 well-known authors of mystery and crime fiction. Each book is colored by its most prominent topic, revealing a broad transition from closed-circle puzzle stories (in magenta and pink), to harder-boiled professional detectives (in green and yellow), to crime thrillers that may or may not have detection at their center (mostly in blue). The sharpness of these divisions is exaggerated in the figure because books aren't really defined exclusively by their *single* largest topic. But tracing the largest topic is a useful way to dramatize the stability of many writers' literary output. The creator of Perry Mason for instance, Erle Stanley Gardner, has 13 books in our corpus, spread over 22 years—all with the same dominant topic. In a few cases, we can see a writer's preferred topic shift systematically across a career (Patricia Highsmith and Chester Himes provide examples). But more commonly, the dividing lines between styles and genres run between authors; topics rise and fall mostly because writers enter or leave the market. It is easy to see why birth year would have explanatory value in a corpus with this structure.

Because many different things are going on at any given moment, books also differ from each other in ways that elude explanation by chronology. In 1969, Mario Puzo published *The Godfather* and Frank Herbert published *Dune Messiah*. Both men were also born in 1920. So the books are indistinguishable as far as dates go. But they are very different in genre and setting, as our topic model certainly reveals. So it shouldn't be surprising that all chronological variables taken together (birth year, publication year, and author's age) can explain only a small part of the differences between books. Averaging across topics and weighting the average by topic size, r^2 for our regression model is only 0.06. This is a small part of overall variance. But the small part of variance explained by time is the important part, of course, for hypotheses about change.

Change within Individual Lives

The regression analysis outlined above asks how period and cohort processes explain change in the corpus as a whole. We supplement this analysis by studying longitudinal change within the careers of individual authors who produced multiple works. To do this, we follow Vaisey and Kiley (2021) and compare a series of structural equation models to evaluate whether an author's use of topics updates durably over time or whether works are best modeled as random departures from an author's settled baseline.

To compare these two processes, we begin with a general model that posits that the topic prevalence, y_{it}, for each work in a triplet—a series of three consecutive works by the same author—is a function of a settled baseline, U_i; some weighting, ρ, of the topic's prevalence in the previous work, y_{it-1}; and work-specific random error, ϵ_{it}. Because the first work in a triplet has no observed prior work, we cannot model it directly and instead include its covariance with U_i to reflect that they share some unobserved causes. This generates the following general model:

$$y_{i3} = \rho y_{i2} + U_i + \epsilon_{i3}, \tag{2}$$

$$y_{i2} = \rho y_{i1} + U_i + \epsilon_{i2}, \tag{3}$$

$$Cov(U_i, y_{i1}) = \tau. \tag{4}$$

We compare the two processes outlined above by testing a constraint on this general model. If we can constrain ρ to 0 without significant decrease in model fit, then deviations from an author's settled baseline can be thought of as random, consistent with an acquired dispositions model. If including a nonzero ρ term in the model improves model fit, we take this as evidence that changes that occur between two works tend to persist to the third, suggesting durable influence from the environment. We estimate models with and without this constraint and compare the Bayesian information criterion (BIC) for the two models, preferring the model with the lower BIC. When the BIC difference between the two models is less than two, we deem the result inconclusive. We compare the settled dispositions model

with the better fitting of two different durable updating models: one that constrains U_i and τ to 0 and one that allows them to be estimated freely.

Because this approach requires three works by the same author, we limit the sample to the 1,714 authors in the sample with three or more works. If authors have more than three works, we look at all sets of three consecutive works, for a total of 6,927 unique triplets. We weight authors by one divided by the number of triplets they have in the sample so that each author contributes the same amount of information to the models.

Structural Equation Results

Models of longitudinal change within authors' careers tell a complex story. Most topics (128 of 200) show some evidence of authors making durable changes over time, although 57 topics show no evidence for any durable updating (Figure 4). The remaining 15 topics do not clearly favor one model or the other.

Patterns of change within the careers of individual authors are clearly related to the separation of cohort and period effects in the century as a whole. In topics where the active updating model is favored, mean δ is 0.496 (weighted, as usual, by r^2 and topic size). In topics where settled dispositions are favored, mean δ is 0.645. A Welch's t test confirms the difference of means between the two groups ($n = 185, t = -4.01, p < 0.001$, two-tailed). Broadly and as a rule, period effects are stronger in topics where individual authors display durable change. This is what we might expect if our intuitive model is that period effects are produced by influences that change writers and cohort effects are strong in topics where writers stay the same throughout their career.

But this broad pattern by no means always holds; there are interesting exceptions. Topic categories that show some of the largest period effects, in particular the "events" category, also show some of the least evidence of updating within authors. One might infer that the strong period effects observed in these topics are sometimes due to temporary shifts in collective attention—associated, for instance, with a war—rather than durable changes in writing styles. Not everyone who wrote about war in the early 1940s became a war novelist forever. "Accidents of transcription" also show little evidence of updating, perhaps for a similar reason: period effects may have been created by temporary changes in printing format rather than changes in writers themselves.

So it is certainly possible to explain strong period effects without evidence of authorial updating. But comparing these two experiments also produces a converse paradox: 69 topics show evidence of updating although they have strong cohort effects ($\delta > 0.5$). It may seem puzzling that cohort effects dominate aspects of writing where authorial practices can change. This puzzle could also be expressed by contrasting the overall results of our two analyses. If cohort variables are more important than period variables overall in the regression analysis, why do our sequence models confirm updating in the majority (64 percent) of cases?
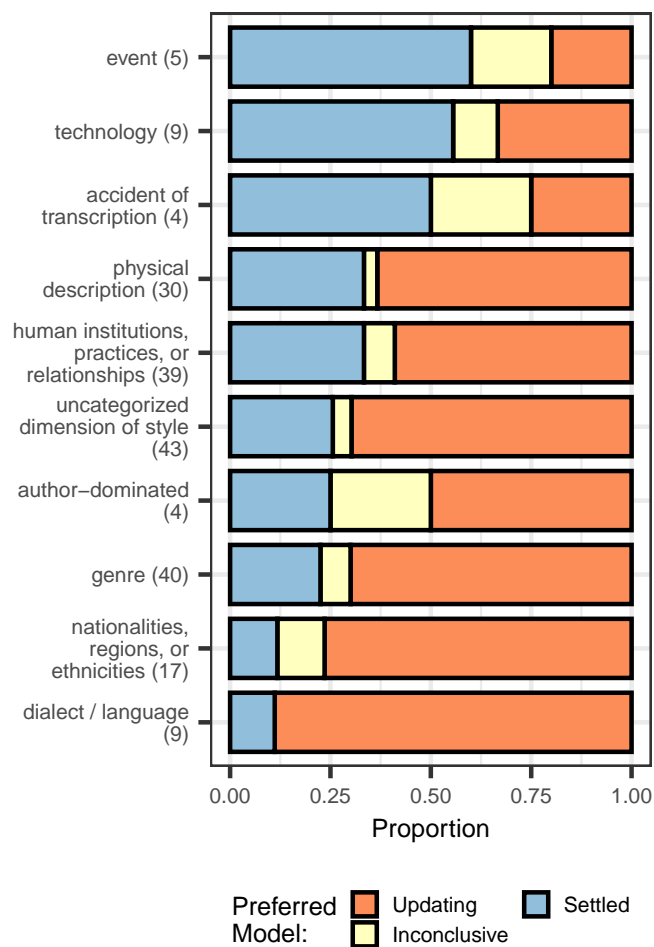
Figure 4: Proportion of topics favoring updating and settled models, by topic category. Total $n = 200$, each estimated on 1,714 unique authors with 6,927 unique triples. Topics favor the updating model when changes in between two consecutive books persist to a third.

Are Cohort Effects Compatible with Individual Change?

As the previous section explained, our regression models and structural equation models mostly align. But they diverge enough to raise an interesting question. Why does evidence for cohort stability appear weaker when we look at individual careers than when we look at the power of birth year to explain the population as a whole?

This is a question our experiment raises rather than one we claim to resolve. But we can offer three interlocking hypotheses, and some tentative evidence, to explain the apparent paradox. First, the structural equation model is posing a binary question: "does active updating happen?" The answer is usually yes. But the regression model is asking a question about the relative magnitude of two types of change. Even when active updating happens, it can still be outweighed by the differences between cohorts.
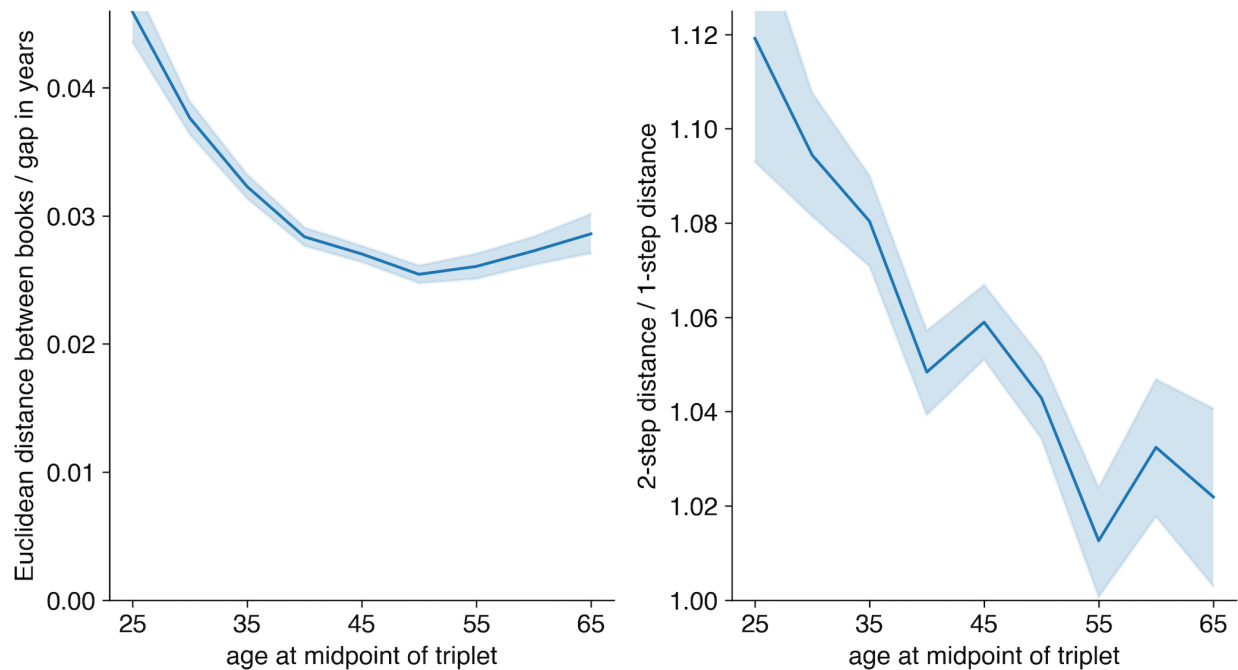
Figure 5: We repeatedly select three books sequential in our representation of an author's career and measure the Euclidean distance between all three pairs in 200-dimensional topic space. The first subplot shows the mean distance from book 1 to book 3, divided by mean gap in years. The second subplot shows the mean distance from book 1 to book 3 as a multiple of the mean distance in the two single time steps (1 to 2 and 2 to 3); this measures the tendency for successive changes to move in the same direction. Distances are measured for 6,927 triplets resampled to give 1,714 authors equal weight; the shaded interval represents a 95 percent confidence interval from bootstrap resampling.

Second, longitudinal changes are not necessarily period effects. If writers in one generation changed in parallel ways across their careers (but changed differently from other generations or at different moments), then the shared pattern of change could appear in our regression model as a cohort effect. And this is not a purely speculative hypothesis. We know that different generations do display different patterns of change because we also see—third—strong evidence that the changes in a writer's career are largest and most persistent in early life.

For instance, if we take the same 6,927 sequences of three books analyzed in the structural equation model and simply measure the Euclidean distance between the first and third books (in 200-dimensional topic space), Figure 5 shows that change across time is significantly more rapid for authors aged 25 to 35. Moreover, a greater proportion of change is durable in early life. If we consider the triangle composed by the three books in each triplet and measure the distance from book 1 to book 3 as a multiple of the average distance between the other two pairs, we see that young writers' careers have a stronger tendency to move in a consistent direction.

The rapidly declining pace of change up to age 35 in Figure 5 is consistent with other evidence for an "impressionable period" in late adolescence and early adulthood (Krosnick and Alwin 1989). It is possible that change may be even more

rapid and more durable for writers younger than 25; we don't have enough books published by writers in their teens and early 20s to draw reliable conclusions. But the evidence we can see at this point already suggests a possible resolution to the tension between our two experiments. Writers do change across their careers. However, cohorts may remain distinct enough to produce a strong cohort effect because writers change most consequentially in early life, which happens for each cohort at a different point in history.

Finally, we should note that this analysis of literary culture is interestingly in tension with previous research using similar methods on rotating panels of the General Social Survey (GSS; 2006 to 2014). In Vaisey and Kiley (2021), only 16 percent of questions in the GSS showed evidence of durable updating, but in our topic model, the updating model is preferred for 64 percent of topics. There are many possible explanations for this difference. The three-book sequences studied here stretch over nine years on average—twice as long as the four-year period covered by three waves of a GSS panel. Durable change might be easier to detect across this longer period. Or there might be a substantive difference between the social domains involved. Opinions reported in surveys are often considered "personal culture," whereas published works reflect "public culture," and change in these domains is theorized to be governed by different processes (Lizardo 2017). Perhaps competition to sell books gives writers of fiction an external incentive to change that is missing in a survey. Differences also might be due to selection. Survey respondents are asked the same questions regardless of whether they have considered an issue between surveys. In contrast, authors might be more likely to write a new book, and get published, when they have something new to say. More research is clearly needed on this question.

Discussion

This analysis has not rejected either of the theories of change it set out to test. Rather, it sets bounds on their relative importance in literary culture. We conclude that differences between cohorts explain slightly more than half of literary change.

But setting a lower bound is sufficient to pose a stark challenge to current historical practice, because the factors that shape cohorts occupy far less than half the pages in literary history. Historians' attention focuses more heavily on the lives of adults: their debates, alliances, scandals, and bad reviews (Stearns 2008). Those things clearly do matter, but they seem to constitute only half the story of cultural change. An equally important part of the story remains, in effect, dark energy.

This is not to say that our results overturn any explicit theory about the relative importance of periods and cohorts in cultural history. Historians are not guided by explicit theory on this topic. Practice varies widely. Some writers have acknowledged that "generational tension" plays an important role in history (Schorske 1998). But it remains true that we have little direct evidence about the developmental processes that define dispositions in a cohort, because "the overwhelming majority of children leave no records of themselves" (Saxton 2008:2). If cohort formation is more than half the story of literary change, current research strategies will need

adjustment. Historians may need to attend more closely, for instance, to books for young readers.

Our study focuses on twentieth-century fiction, and one case study would not by itself support broader inferences about culture. But fiction is far from the only domain where strong cohort effects have been observed. On the contrary, literary studies is one of the last remaining strongholds of the approach that explains cultural change mainly by dividing history into periods. Phonology and grammar are already known to change through cohort succession (Meisel et al. 2013). And it has been clear for several decades that political attitudes tend to persist across a life: even relatively popular discussions of political behavior are now organized generationally (Miller and Shanks 1996). But novels are still more commonly interpreted as typifying a period—or even a single decade.

Periodizing impulses are especially strong in discussion of recent literary history, because twentieth-century debates and crises still loom large enough in public memory to suggest a finely divided "hot" chronology (Lévi-Strauss 1962:259). Critics speak not just about "modernism" and "postmodernism" but about "Depression-era fiction," the "paperback revolution," and "the 9/11 novel" (Anker 2011; Mercer 2011). Historical crises do, of course, leave traces in stories. However, we find that even in twentieth-century fiction—where critical tradition seems to predict periodizing ruptures—cohort effects are the larger part of cultural change. We can't necessarily infer that the same thing will hold true in all domains of culture, but at this point there is mounting evidence that it is true more often than not.

Our results dovetail with several recent studies suggesting that literary change is a more continuous process than historians once assumed. Researchers have struggled to find the breaks often described in literary history: decade-sized "turns," or intervals of rapid transformation that separate stable genres and periods (Beausang 2021; Šeļa, Plecháč, and Lassche 2021). What they find instead are trends with massive momentum, proceeding in the same direction for a century or more (Underwood 2019). This level of continuity would become easier to understand if literary change depended largely on cohort replacement—a process that is necessarily gradual because cohorts always overlap.

Beyond literature, our conclusion is compatible with an emphasis on "cultural transmission" that organizes much research on cultural evolution (Mesoudi and Whiten 2008). But a cohort-centered model will pose a challenge for the common experimental strategy of studying transmission between adult participants within a relatively short time window (Fay et al. 2019; Lucas et al. 2020). Since cohort formation takes decades, the transmission of cultural practices may often be a long-term developmental process rather than a rapid imitative one. Where this is true, researchers will need to question the model of cultural evolution that envisions cultural artefacts themselves as nodes in a branching tree of influence and descent. Instead of modeling selective pressures on "the populations of poetry and fiction," we may need to focus on populations of people and the slow processes that shape them (Moretti and Sobchuk 2019:112).

But these are large questions; they certainly have not been resolved by one experiment. On the contrary, our research opens up several new avenues of inquiry. We found that topics varied widely in their strength of association with period

or with cohort variables. What factors explain that variation? The answer might help researchers move beyond describing the pace of cultural change and toward understanding its causes. We also found a striking decline in the pace of individual change as authors grew older. The sources of change and the reasons it tends to decline with age together compose a research question with clear social importance.

Appendix: Detailed Methods

Data Construction

To identify works of fiction in the period 1880 to 1999 we relied on the NovelTM Datasets for English-Language Fiction (Underwood et al. 2020). But the metadata in that data set needed to be enriched for our experiment. Birth years were available for many but not all authors, and the metadata provided an estimate rather than a confirmed first-publication date. This estimate is usually close to the date of first publication, but in eight percent of volumes the two figures diverged by more than six years.

Since publication year is a central variable for our experiment, we needed a better estimate. To achieve that, we tagged books manually, collated several lists manually corrected by other researchers, and also matched the NovelTM records against the U.S. Copyright Registry, digitized by the New York Public Library (Cram et al. 2019). This gave us a subset of 10,830 volumes where we had strong confidence in both publication date and author's date of birth. All conclusions above are based on this subset. Of these, a subset of 5,572 were also manually identified as books by authors who resided in the United States for most of their careers; the others are by authors of unconfirmed nationality, although the majority are also in practice Americans.

This selection process was likely to favor moderately well-known writers, since date of birth can be impossible to identify when a writer is truly obscure. But that selection bias matches one implicit in the theories of literary culture we are attempting to evaluate—which usually take aim at an even smaller sample of clearly prominent authors.

Because the granularity of a topic model tends to coarsen at the edge of the timeline, we topic-modeled 120 years from 1880 to 1999 but focused our analysis on the central century: 1890 to 1989. We also kept the number of words per decade constant across the model, although the subset of books with confirmed birth year and publication date was largest in the center of the period.

Topic Model

We downloaded page-level word frequencies for all 29,351 volumes from HathiTrust Digital Library (Capitanu et al. 2016). To reduce the influence of front matter, biographical introductions, and advertisements, we discarded the first 15 percent and last five percent of pages in each volume. Then we divided volumes into segments of approximately 10,000 words and topic-modeled the 154,883 segments.

We excluded only a short list of 26 common words, Arabic numerals, personal names, and very rare words (that occurred in fewer than 35 authors). Except for those exclusions, every word in the corpus was assigned to some topic. We used MALLET for the modeling (McCallum 2002).

Topic modeling is an unsupervised process governed by several hyperparameters. Of these, the one that attracts most debate is k—the number of topics in the model. Other hyperparameters can be optimized using an inference process internal to the model (Wallach, Mimno, and McCallum 2009). But there are several different ways to optimize k. Many researchers argue that purely internal criteria are never sufficient to determine a "correct" number of topics. Different settings can be equally correct, so the choice must be guided instead by the objectives of the research and by "qualitative human judgment" (Maier et al. 2018).

The importance of this decision also varies with the purpose of the model. If researchers expect the boundaries of topics to correspond to objective divisions in the corpus, then it becomes important to demonstrate that a corpus has been divided into the right number of parts. But other researchers use topic modeling simply to reduce the dimensionality of evidence in a context-sensitive way. If that's the purpose of modeling, then the boundaries between topics are arbitrary conveniences, and many different settings of k could work equally well. Recent research on the French Revolution by Barron et al. (2018) provides one example of a project using topic modeling in this way; our article is another.

Taking this view of the problem, we don't claim to have discovered an "optimal" number of topics. Our goal is rather to find a level of granularity that supports the larger goals of our experiment—a level of granularity, in other words, roughly comparable to the way human beings already discuss literary change.

Since we wanted to reason about a scale of analysis typically practiced by human readers, we selected a model whose granularity matched the granularity of subject and genre categories for the same volumes identified by librarians and by contemporaneous readers in the *Book Review Digest* (Wilson 1906–1999). We supplemented library metadata with the *Digest* because genre labels assigned by contemporary librarians are sparse in the earlier twentieth century.

Then we trained different models and measured their correspondence with our ground truth using two different metrics. Adjusted mutual information requires a hard partitioning, so to use this metric we assigned each document to its most prominent topic; this corresponded to the discrete character of our ground truth. But a topic model is fundamentally a fuzzy partition, so we also used a version of the adjusted Rand index adapted by Brouwer (2009) for fuzzy partitioning schemes. The results are visualized in Figure 6. Since our goal is to maximize correspondence with ground truth on both axes, models near the upper right corner are most desirable. There is a trade-off between goals, and any setting of k between 100 or 400 might be justifiable; we erred on the low side (200) because we were concerned that higher settings might produce author-specific topics that would be dominated by cohort effects for circular and unenlightening reasons. (For the same reason, we excluded topics with significant concentration on one author from our calculation of overall δ.)
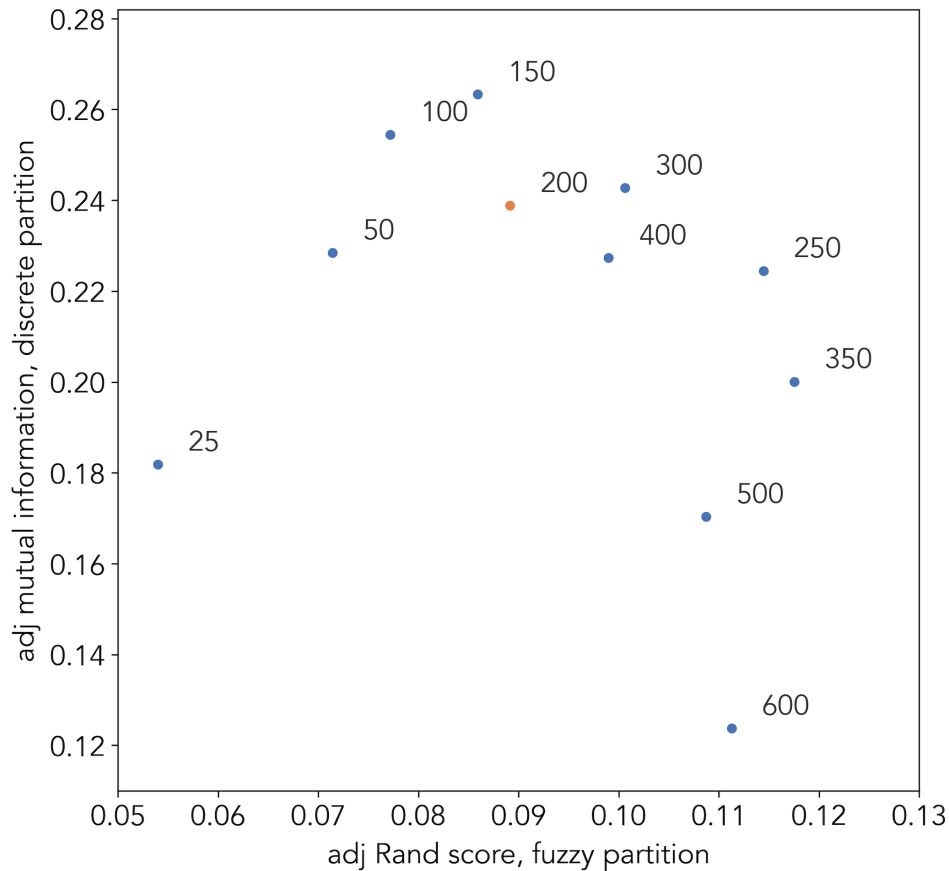
Figure 6: Strength of alignment between topic models and a partitioning of the same volumes in human ground truth. Both "adjusted mutual information" and "adjusted Rand score" are higher when volumes in the same human category are also characterized similarly by the model.

Settings of *k* between 200 and 400 are already common for large corpora and recommended, for instance, by the developers of MALLET (McCallum 2002). The process we describe above merely confirms that customary practices are empirically justified in the context of our project and data.

Topic Categories

Topics can be grouped into categories only very loosely. Topics with many common words are often opaque; for this reason, we included "uncategorized" as one of our 10 categories. Large, diffuse topics were often assigned there. Even when the words themselves are unambiguous, it is not easy to draw a line between physical and social concepts. Topic 91 below is clearly about gardens and plants. But should we call a garden an "institution," a "technology," or "an aspect of physical description"? Since uncultivated landscapes may also be included in topic 91, we went with "physical description," but the line is not crisp.

Table 1: Examples of topic categories

Topic label	Category	Keywords	Top author	Peak dates
0. War, mostly WWII	event	war german french germans	Upton Sinclair 3.15%	1918–1957
3. Late 20c US pol. thrillers	genre	american people president our	Tom Clancy 2.86%	1976–1991
15. French language	dialects/languages	de madame le la monsieur	Victor Hugo 3.31%	1899–1942
45. Wodehouse	author-dominated	butler man thing yes say just	P.G. Wodehouse 45.28%	1924–1974
91. Gardens and plants	physical description	garden flowers green trees	Beverley Nichols 0.35%	1900–1941
93. British peerage	institutions/practices	lady lord earl madam	Angela Thirkell 0.85%	1899–1927
122. Aeronautics	technology	air plane pilot flying flight	Edward Ellsberg 1.27%	1942–1974
128. Jews and Judaism	nationalities/ethnicities	jews jew jewish rabbi israel	Scholem Asch 6.63%	1933–1981
143. Dialogue?	uncategorized	am know see must shall	Unknown 0.25%	1903–1967
176. ads at the back	accident of transcription	v cloth crown j c w edition h net	Unknown 1.64%	1887–1904

The coding guide we used for categorization is available in our public code repository. In categorizing topics we considered not only the information in Table 1 but also statistical information about the distribution of topics across documents. We found that topics associated with "genres," for instance, tended to concentrate in a relatively small number of books where they were relatively dominant.

Note that the size of topics varies widely. In Table 1, the uncategorized topic 143 contains 1.22 percent of the tokens in the corpus, roughly 11 times as many as topic 15, which is composed largely of French words.

"Accidents of transcription" include changes in publishing format and orthography as well as OCR errors. The example above comes from a period when publishers included catalogs at the back of many books; although we trim the last five percent of pages to minimize paratext, words like "cloth" and "edition" reveal that some pages of publishers' advertisements got through. Varying hyphenation practices also make word segmentation errors ("don re tion con") concentrate in some texts.

We excluded these accidents of publication, as well as author-dominated topics, from our calculation of overall δ. (We defined "author-dominated topics" as those with nine percent or more of tokens coming from a single author.) Excluding strongly author-focused topics is a debatable choice; one could argue that the model's tendency to align topics with authors is a meaningful sign of authorial stability and part of the phenomenon we set out to measure. But it seemed simplest to err on the side of conservatism, at the risk of slightly underestimating δ.

Data Availability

The raw word frequencies for original texts used in this study are available from HathiTrust Research Center and can be downloaded using our metadata and code. The topic model used in the study is archived in Zenodo: https://doi.org/10.5281/zenodo.5515507. The code and metadata used in the study are also archived separately in Zenodo: https://doi.org/10.5281/zenodo.5573232.

For fuller description of models, alternate models, and a detailed table annotating all 200 topics, see the online supplement to this article.

Notes

[1] Neural language models will eventually improve representations of literary history. But at the moment there is little evidence that they represent macroscopic change more faithfully than topic models do. BERT, for instance, is certainly better at reasoning about sentences and paragraphs and even about the sentiment of short documents. But our initial experiments have not shown that it significantly improves predictions about the differences between books.

References

Alwin, Duane F., and Jon A. Krosnick. 1991. "Aging, Cohorts, and the Stability of Sociopolitical Orientations over the Life Span." *American Journal of Sociology* 97(1):169–95. https://doi.org/10.1086/229744.

Andersen, Henning. 1973. "Abductive and Deductive Change." *Language* 49(4):765–93. https://doi.org/10.2307/412063.

Anker, Elizabeth S. 2011. "Allegories of Falling and the 9/11 Novel." *American Literary History* 23(3):463–82. https://doi.org/10.1093/alh/ajr017.

Barron, Alexander T. J., Jenny Huang, Rebecca L. Spang, and Simon DeDeo. 2018. "Individuals, Institutions, and Innovation in the Debates of the French Revolution." *Proceedings of the National Academy of Sciences* 115(18):4607–12. https://doi.org/10.1073/pnas.1717729115.

Beausang, Chris. 2021. "Diachronic Delta: A Computational Method for Analysing Periods of Accelerated Change in Literary Datasets." *Digital Scholarship in the Humanities* advance article, published November 2, 2021. https://doi.org/10.1093/llc/fqab041.

Bell, Andrew J., and Kelvyn Jones. 2013. "The Impossibility of Separating Age, Period and Cohort Effects." *Social Science & Medicine* 93:163–65. https://doi.org/10.1016/j.socscimed.2013.04.029.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3:993–1022. https://dl.acm.org/doi/10.5555/944919.944937.

Bourdieu, Pierre. 1990. *The Logic of Practice*. Stanford University Press.

Brouwer, Roelof K. 2009. "Extending the Rand, Adjusted Rand and Jaccard Indices to Fuzzy Partitions." *Journal of Intelligent Information Systems* 32:213–35. https://doi.org/10.1007/s10844-008-0054-7.

Capitanu, Boris, Ted Underwood, Peter Organisciak, Timothy Cole, Maria Janina Sarol, and J. Stephen Downie. 2016. "The HathiTrust Research Center Extracted Feature Dataset." HathiTrust Research Center. http://doi.org/10.13012/J8X63JT3.

Cram, Greg, Sean Redmond, Josh Hadro, and Cassandra Flores. 2019. "Catalog of Copyright Entries Project." New York Public Library. https://github.com/NYPL/catalog_of_copyright_entries_project.

DiMaggio, Paul. 1997. "Culture and Cognition." *Annual Review of Sociology* 23:263–87. https://doi.org/10.1146/annurev.soc.23.1.263.

Fay, Nicolas, Naomi De Kleine, Bradley Walker, and Christine A. Caldwell. 2019. "Increasing Population Size Can Inhibit Cumulative Cultural Evolution." *Proceedings of the National Academy of Sciences* 116(14):6726–31. https://doi.org/10.1073/pnas.1811413116.

James, Pearl. 2013. *The New Death: American Modernism and World War I*. University of Virginia Press.

Krosnick, Jon A., and Duane F. Alwin. 1989. "Aging and Susceptibility to Attitude Change." *Journal of Personality and Social Psychology* 57(3):416–25. https://doi.org/10.1037/0022-3514.57.3.416.

Lambert, Ben, Georgios Kontonatsios, Matthias Mauch, Theodore Kokkoris, Matthew Jockers, Sophia Ananiadou, and Armand M. Leroi. 2020. "The Pace of Modern Culture." *Nature Human Behaviour* 4:352–60. https://doi.org/10.1038/s41562-019-0802-4.

Lévi-Strauss, Claude. 1962. *The Savage Mind*. University of Chicago Press.

Lizardo, Omar. 2017. "Improving Cultural Analysis: Considering Personal Culture in Its Declarative and Nondeclarative Modes." *American Sociological Review* 82(1):88–115. https://doi.org/10.1177/0003122416675175.

Lucas, Amanda J., Michael Kings, Devi Whittle, Emma Davey, Francesca Happé, Christine A. Caldwell, and Alex Thornton. 2020. "The Value of Teaching Increases with Tool Complexity in Cumulative Cultural Evolution." *Proceedings of the Royal Society B: Biological Sciences* 287(1939):20201885. https://doi.org/10.1098/rspb.2020.1885.

Maier, Daniel, A. Waldherr, P. Miltner, G. Wiedemann, Niekler A., A. Keinert, B. Pfetsch, G. Heyer, U. Reber, T. Häussler, H. Schmid-Petri, and S. Adam. 2018. "Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology." *Communication Methods and Measures* 12(2–3):93–118. https://doi.org/10.1080/19312458.2018.1430754.

Mason, Karen Oppenheim, William M. Mason, H. H. Winsborough, and Poole W. Kenneth. 1973. "Some Methodological Issues in Cohort Analysis of Archival Data." *American Sociological Review* 38(2):242—-58. https://doi.org/10.2307/2094398.

McCallum, Andrew Kachites. 2002. "MALLET: A Machine Learning for Language Toolkit." http://mallet.cs.umass.edu.

Meisel, Jurgen M., Martin Elsig, and Esther Rinke. 2013. *Language Acquisition and Change: A Morphosyntactic Perspective*. Edinburgh University Press.

Mercer, Ben. 2011. "The Paperback Revolution: Mass-Circulation Books and the Cultural Origins of 1968 in Western Europe." *Journal of the History of Ideas* 72(4):613–36. https://doi.org/10.1353/jhi.2011.0032.

Mesoudi, Alex, and Andrew Whiten. 2008. "The Multiple Roles of Cultural Transmission Experiments in Understanding Human Cultural Evolution." *Philosophical Transactions of the Royal Society B: Biological Sciences* 363(1509):3489–501. https://doi.org/10.1098/rstb.2008.0129.

Miller, Warren E., and J. Merrill Shanks. 1996. *The New American Voter*. Cambridge, MA: Harvard University Press.

Moretti, Franco, and Oleg Sobchuk. 2019. "Hidden in Plain Sight: Data Visualization in the Humanities." *New Left Review* 118:86–115.

Nelson, Laura K. 2017. "Computational Grounded Theory: A Methodological Framework." *Sociological Methods & Research* 49(1):3–42. https://doi.org/10.1177/0049124117729703.

Ramazani, Jahan, and Jon Stallworthy. 2012. "Introduction." In *The Norton Anthology of English Literature, Volume F: The Twentieth Century and After*, pp. 1887–910. New York: W. W. Norton, 9th edition.

Ryder, Norman B. 1965. "The Cohort as a Concept in the Study of Social Change." *American Sociological Review* 30(6):843–61. https://doi.org/10.2307/2090964.

Saxton, Martha. 2008. "Introduction." *Journal of the History of Childhood and Youth* 1(1):1–3. https://muse.jhu.edu/article/233325.

Schorske, Carl E. 1998. "Generational Tension and Cultural Change." In *Thinking with History: Explorations in the Passage to Modernism*, edited by Carl E. Schorske, pp. 141–56. Princeton University Press.

Sears, David O., and Carolyn L. Funk. 1999. "Evidence of the Long-Term Persistence of Adults' Political Predispositions." *Journal of Politics* 61(1):1–28. https://doi.org/10.2307/2647773.

Šela, Artjoms, Petr Plecháč, and Alie Lassche. 2021. "Semantics of European Poetry Is Shaped by Conservative Forces: The Relationship between Poetic Meter and Meaning in Accentual-Syllabic Verse." Preprint, arXiV:2109.07148 [cs.CL]. https://doi.org/10.48550/arXiv.2109.07148.

Sherry, Vincent. 2003. *The Great War and the Language of Modernism*. Oxford University Press.

Stearns, Peter N. 2008. "Challenges in the History of Childhood." *Journal of the History of Childhood and Youth* 1(1):32–42. https://muse.jhu.edu/article/233326.

Underwood, Ted. 2019. *Distant Horizons*. University of Chicago Press.

Underwood, Ted, Kevin Kiley, Wenyi Shang, and Stephen Vaisey. 2021. "Period and Cohort Effects in English-Language Fiction, 1890–1989." Preregistration on Open Science Framework Registries, July 9, 2021. https://doi.org/10.17605/OSF.IO/4E2K7.

Underwood, Ted, Patrick Kimutis, and Jessica Witte. 2020. "NovelTM Datasets for English-Language Fiction, 1700–2009." *Journal of Cultural Analytics* 5(2):13147. https://doi.org/10.22148/001c.13147.

Vaisey, Stephen, and Kevin Kiley. 2021. "A Model-Based Method for Detecting Persistent Cultural Change Using Panel Data." *Sociological Science* 8(5):83–95. https://doi.org/10.15195/v8.a5.

Vaisey, Stephen, and Omar Lizardo. 2016. "Cultural Fragmentation or Acquired Dispositions? A New Approach to Accounting for Patterns of Cultural Change." *Socius* 2:1–15. https://doi.org/10.1177/2378023116669726.

van Cranenburgh, Andreas, Karina van Dalen-Oskam, and Joris van Zundert. 2019. "Vector Space Explorations of Literary Language." *Language Resources and Evaluation* 53:625–50. https://doi.org/10.1007/s10579-018-09442-4.

Wallach, Hanna, David Mimno, and Andrew McCallum. 2009. "Rethinking LDA: Why Priors Matter." In *Advances in Neural Information Processing Systems 22*, edited by Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta. Red Hook, NY: Curran Associates, Inc.

Wilson, H. W. 1906–1999. *The Book Review Digest*. Bronx, NY: H. W. Wilson.

Ted Underwood: School of Information Sciences, University of Illinois, Champaign-Urbana. E-mail: tunder@illinois.edu.

Kevin Kiley: Department of Sociology and Criminology, University of Iowa.
E-mail: kevin-kiley@uiowa.edu.

Wenyi Shang: School of Information Sciences, University of Illinois, Champaign-Urbana.
E-mail: wenyis3@illinois.edu.

Stephen Vaisey: Department of Sociology, Duke University.
E-mail: stephen.vaisey @duke.edu.