Have any virtues become anomalous in popular literature in Western societies?

Technical report

Date: 20 November 2026

Question framing: This is an open question for the evidence. This report does not assume that the answer is yes or no.

In brief

The question behind this brief is whether any virtues have become unusual, or have come to be treated as flaws, in the popular fiction of Western societies. The evidence so far answers part of that question. It shows how often virtue words appear in books. How popular novels portray the people who practise those virtues has so far been tested only in two small pilots.

The clearest finding concerns words. In a 2012 study, Pelin and Selin Kesebir tracked 50 virtue words, such as honesty, patience, kindness, and courage, in American books and found that most became less frequent over the twentieth century. Their results faced one obvious objection: the Google Books collection took in more and more scientific writing, which could push ordinary words down. We reran their exact word list on the fiction-only part of the collection, where that problem largely disappears. The result held, and was slightly stronger. Between 1901 and 2000, 41 of the 50 words fell in fiction, and the typical word was used about half as often by the end of the century. Run on the Kesebirs' own collection, the same method reproduced their published counts exactly. Ordinary words drift as well, so each virtue word was also compared with ordinary words that were about as common around 1900. Those ordinary words fell too, but much less. Allowing for that, the typical virtue word still fell by about two fifths in fiction and by about a third in books generally. For the ten general moral terms, such as "virtue" and "conscience", the fall cannot be told apart from ordinary drift.

We then looked at seven groups of words tied to virtues that are often said to be viewed differently today: duty to parents and family; obedience; religious devotion; humility and modesty; self-control and composure; sexual restraint and chastity; and putting others before oneself. Every group became less frequent between the 1900s and the 1980s and early 1990s, in fiction and in books generally. By then each group was used at between roughly a third and seven tenths of its early-century rate, and the fall remains after allowing for the drift of ordinary words. In books generally, most groups rose again by the 2000s while fiction stayed roughly flat, but most of that difference disappears once ordinary-word drift is allowed for. Two groups are in effect single words: "duty" makes up about 95% of the family-duty group and "devotion" about 85% of the self-sacrifice group. Counting each word equally, the fall holds for obedience, religious devotion, humility, and self-control in both collections. It does not hold for self-sacrifice: without "devotion", that group shows no fall, because "selfless" rose many times over while "self-sacrifice" and "self-denial" fell. In fiction, the falls for family duty and chastity rest mainly on "duty" and "purity".

Older research points the same way. In nearly 3,000 British novels of the nineteenth century, words for abstract values, including moral judgment and social restraint, together fell by more than half. Across the twentieth century, some broad moral vocabulary dropped until about 1980 and then rose again, and words such as "obliged" and "obedience" grew rarer from 1800 to 2000.

What was missing was the direct test. No study we found has looked at bestselling novels at two or more dates and counted how often characters practise a given virtue, or whether the story presents that virtue as admirable or as a flaw, against less popular novels from the same years. We ran two small pilots of that test. The first took ten bestsellers from the 1900s and ten from the 1920s, paired each with a less popular novel from the same years, and coded short passages from the full texts. The second compared the 1920s with the 1990s using published plot summaries and reviews, which is a weaker measure. Neither pilot found a reliable sign that any of the seven virtues became rarer in bestsellers than in the less popular novels, or more often treated as a flaw. One result in the first pilot, for duty to parents, did not survive allowing for seven separate tests. The pilots are too small to rule such changes out: with ten books in each group, only very large shifts would show. The result is therefore inconclusive, not a finding of no change.

The pilots also showed what a full test would need. Short passages rarely reveal how a story judges a character: in the first pilot, the passages showed no judgment for four in five of the characters coded. Summaries reveal the judgment but mention few virtues at all, and the two methods seldom picked out the same virtue in the same book. A full test would need several times as many books, longer readings, and independent human coders.

Some cautions apply. The book collections studied are library holdings, not lists of what people actually read. Fewer uses of a word do not show that a virtue is portrayed differently. Many of these words have other meanings. Ordinary words drift too; allowing for that shrinks the virtue-word decline but does not remove it, and the matching is on frequency only. The seven word lists came from an outside chart of unknown origin, and the contrast between fiction and other books after the 1990s was noticed in the data rather than predicted in advance. The pilots were coded by one AI system, which could not be kept unaware of each book's period, so their agreement figures overstate reliability.

Executive summary

- **The question has now been tested at the unit it needs, but only in two small pilots (section 2.8, November 17, 2026).** Bestsellers against matched, less popular novels, 10 works per cell: 1900s to 1920s from full-text extracts, and 1920s to 1990s from summaries and reviews. Both are inconclusive for H1 and H1F: no difference was detected, but every interval includes zero and the samples are too small to rule out moderate effects. Pooled rarity differences in differences were +0.07 (95% interval -0.13 to 0.29) and +0.09 (-0.01 to 0.19), the opposite sign to H1; flaw-framing differences were +0.12 (-0.04 to 0.29) and -0.28 (-0.88 to 0.33). The pooled intervals rule out only an average extra fall in bestsellers of more than 0.13 (Pilot A) or 0.01 (Pilot B, on base shares of 0.01 to 0.13) in the share of works, and a pooled rise in flaw framing above 0.29 (Pilot A); per-virtue intervals leave large effects open. Underpowered, one AI coder, agreement inflated. The search (Appendix B) found no published study of this kind for any virtue.

- **New run on Google Books English Fiction (2012), November 16, 2026:** with Kesebir and Kesebir's exact term lists, 41 of 50 virtue words and 8 of 10 general moral terms fell significantly from 1901 to 2000; the median virtue word fell by about half. All English books show a similar pattern (38 of 50; median fall 47%), so the decline is not mainly an artifact of scientific texts. The same script reproduces Kesebir and Kesebir's published counts exactly on their own corpus. Against frequency-matched ordinary words (new, November 18), the median virtue word still fell 41% in fiction and 33% to 37% in all books, intervals excluding zero; for the 10 general terms the adjusted fall is not distinguishable from drift.
- **New run on term groups for the seven candidate shifts (lexical only, not framing):** every group (filial piety, obedience, religious piety, humility, restraint, chastity, self-sacrifice) fell from 1900-09 to 1983-94 in both fiction and all English books, to between 36% and 71% of its starting level. All books then rose again in 2000-08 in most groups, while fiction stayed roughly flat. The falls remain after adjustment for ordinary-word drift, but the post-1990s contrast between fiction and all books largely disappears. C1 and C7 are effectively single-word measures (*duty*, *devotion*). With equal weights, or without the most common word, the fall holds for C2 to C5 in both corpora, but not for C7, and in fiction C1 and C6 are then not distinguishable from drift (section 2.4).
- **Earlier term studies agree in direction, with one reversal:** moral word fields fell in nineteenth-century British novels (Heuser and Le-Khac 2012); general-morality and purity terms fell until about 1980 and then rose (Wheeler et al. 2019); *obliged* and *obedience* fell from 1800 to 2000 (Greenfield 2013).
- **A full test is practical.** Character-level coding of about 104,000 works of fiction, 1780 to 2007, with a check on *Publishers Weekly* bestsellers, has been done for gender (Underwood, Bamman and Lee 2018), and annual US bestseller lists for 1895 to 1975 are compiled (Hackett and Burke 1977).
- **Appendix A is unsourced:** seven candidate framing shifts and a coding scheme, offered as hypotheses only. The pilots neither confirmed nor refuted any of them.

1. Working definitions

Anomalous virtue (the standard definition used in all CAV briefs). A virtue that departs from its setting's norm (unusual, unexpected, or contradictory there), or that does not carry over into another moral vocabulary without something being lost. Two readings are tested.

- **Rarity measure (rarity reading).** Share of *works* in a dated sample in which a principal character practises virtue V; the denominator is all works in the sample.
- **Framing measure (remainder reading).** Share of *practising principal characters* whose practice the work frames as a flaw (the arc punishes or corrects it, or the narrator or other characters treat it as a fault); the denominator is all principal characters who practise V. The endorsed share uses the same denominator; *mixed* and *not evaluated* are reported separately.
- **Frequent but pathologized virtues** count as anomalous under the remainder reading, so a virtue can pass the framing test while failing the rarity test.

Main test (relative). Compare the corpus of interest with a justified match (same period, language, and market; different popularity tier, genre, or audience) using the same definition and records. A virtue *became* anomalous if, between two dates, its rarity share fell or its flaw-framed share rose by more than in the matched corpus (a difference in differences whose confidence interval excludes zero). Without matched data, the question is untested.

No 50% line. Almost no single virtue is depicted in a majority of popular works or characters at any date, so a below-50% rule would mark nearly every virtue anomalous at all times and separate nothing. Only the relative test is used.

Popular literature. Works meeting a documented popularity standard in a named market at the time (a bestseller list, or a sales or circulation threshold). A library with one copy of each book is not a popularity sample. **Western societies** means only the named markets in each source. **"Become"** requires the same kind of record at two or more dates.

Terms versus portrayals. A *virtue-term* is a word from a fixed list; its frequency counts the string. A *virtue-portrayal* is a character or work shown practising or endorsing a virtue, and requires coding of characters or plots. The term studies below measure only the first.

2. Evidence

2.1 What word frequency can show

Pechenick, Danforth and Dodds (2015) show that Google Books is "in effect a library, containing one of each book," and that scientific texts grew as a share of English data through the twentieth century. Only the 2012 English Fiction set "is not heavily affected by professional texts."

2.2 Term studies of English-language books

- **Kesebir and Kesebir (2012),** American books, 1901 to 2000. Of 10 general moral nouns (*character, conscience, decency, dignity, ethics, morality, rectitude, righteousness, uprightness, virtue*), eight fell significantly, *ethics* rose, and *morality* had no significant trend. Of 50 rated virtue words, 37 (74%) fell, 8 rose, and 5 had no significant trend.
- **Wheeler, McGrath and Haslam (2019),** 304 moral terms in the English (2012) set, 1900 to 2007. General-morality and purity terms fell until about 1980 and then "rebounded sharply"; harm terms rose after 1980. The fiction subset was not used.
- **Greenfield (2013),** American and British books, 1800 to 2000. *Obliged* and *duty* fell while *choose* and *decision* rose; *obedience*, *authority*, *belong*, and *pray* fell. These bear on C1 and C2 (Appendix A), at term level.
- **Twenge, Campbell and Gentile (2012),** American books, 1960 to 2008. Individualistic words and phrases (*unique*, *self*, "I am special") rose, controlling for communal words. This bears on C4 and C7, at term level.

2.3 New run: Kesebir and Kesebir's terms in the English Fiction subset

Method. Kesebir and Kesebir's exact 10 general terms and 50 virtue words (their Tables 1 and 2), queried through the Ngram Viewer JSON endpoint on November 16, 2026, 5:48 to 5:58 PM MT. Years 1901 to 2000, smoothing 0 (raw annual values), lowercase, case-sensitive. Per word: Pearson correlation of annual frequency with year ($n = 100$; $p < .05$ when $|r| > 0.197$) and percent change from the 1901-1910 mean to the 1991-2000 mean. Corpora: English Fiction 2012 (target); English 2012 and American English 2012 (all books, same version); American English 2009 (the version available when Kesebir and Kesebir wrote, as a replication check).

Corpus	General terms (10): down / up / n.s.	Virtue words (50): down / up / n.s.	Median change per virtue word
American English 2009 (replication)	8 / 1 / 1	37 / 8 / 5	-46%
American English 2012	9 / 1 / 0	37 / 10 / 3	-46%
English 2012 (all books)	9 / 1 / 0	38 / 10 / 2	-47%
English Fiction 2012	8 / 1 / 1	41 / 5 / 4	-50%

Results.

- The replication row matches the published counts exactly, for both term sets.
- In fiction, 41 of 50 virtue words (82%) fell. Only *integrity*, *tolerance*, *selflessness*, *dependability*, and *reliability* rose, three of which Kesebir and Kesebir tie partly to technical senses. *Compassion*, *respect*, *fairness*, *discipline*, and *determination* rose in all English books but not in fiction.
- In fiction, *humility* fell 57%, *humbleness* 71%, *modesty* 52%, *sacrifice* 50%, *faith* 50%, and *purity* 51% (decade means). Against all English books, 30 of 50 words fell more, or rose less, in fiction.
- This compares genres within one library-like corpus, not popular with less popular works.

Drift baseline (new, November 18, 2026). Each target word was compared with five ordinary words of similar 1901-1910 frequency in the same corpus, measured by the same script settings. The rule and seed (20261118) were written before any control data were fetched (`litbrief-tests/drift/drift_prereg.md`). Candidate pool: a seeded random sample of 2,400 nouns and adjectives from the 40,000 most common English words in the wordfreq list (Speer 2022), after removing proper names (WordNet; Miller 1995), every word in the VADER sentiment lexicon of about 7,500 rated evaluative words (Hutto and Gilbert 2014), the target words and their derived forms, and a fixed list of moral and religious words. For each target, five controls were drawn at random from the 20 unused candidates of the same part of speech nearest in 1901-1910 frequency (median gap 2%; largest a factor of 2.5). Drift-adjusted change is the word's ratio of 1991-2000 to 1901-1910 means, divided by the median ratio of its five controls, minus 1. Intervals: 10,000 bootstrap draws resampling words, and controls within words. Medians use the same convention as the table above. Queried November 18, 2026, 6:54 to 7:18 PM MT.

Corpus	Virtue words, raw median change	Matched controls, median change (middle half)	Virtue words, drift-adjusted median (95% interval)	Virtue words falling more than their controls	General terms, drift-adjusted median (95% interval)
English Fiction 2012	-50%	-8% (-46% to +36%)	-41% (-51% to -26%)	41 of 50	-20% (-57% to +56%)
American English 2012	-46%	-10% (-45% to +43%)	-37% (-47% to -15%)	38 of 50	-40% (-64% to +22%)
English 2012 (all books)	-47%	-16% (-50% to +58%)	-33% (-50% to -16%)	37 of 50	-48% (-63% to +31%)

- Matched ordinary words fell too, but less, and with a wide spread. Of the 250 virtue-word controls, 48% to 55% had a significant downward trend, against 74% to 82% of the virtue words. Across the whole sample (about 2,100 words per corpus) the median change was +6% to +14%, so matching on starting frequency matters: words common around 1900 fell more than words in general.
- After adjustment, the median virtue word still fell 41% in fiction and 37% and 33% in all books, with intervals excluding zero. On a log scale the adjustment removes about a quarter (fiction, American English 2012) to two fifths (English 2012) of the raw decline.
- For the 10 general terms, the adjusted decline is not distinguishable from drift: medians -20% to -48%, every interval includes zero (7 or 8 of 10 terms fell more than their controls).
- A post hoc check dropped 48 selected controls with moral, religious, or evaluative content that the filters missed (for example *unholy*, *martyrs*, *malice*). The adjusted virtue medians did not change.

2.4 New run: term groups for C1-C7 (lexical test only)

What this tests. Whether word groups tied to the Appendix A candidates fell in frequency, in fiction and in all English books. It tests lexical frequency only, not framing: it cannot show whether a virtue is endorsed or treated as a flaw.

Method. The seven term lists come from a chart supplied for review on November 16, 2026 (shown under each panel); its author and selection method are not documented, and the lists are not Kesebir and Kesebir's. Every value below was recomputed here on November 16, 2026, 7:27 PM MT, from the Ngram Viewer JSON endpoint: corpora en-fiction-2012 and en-2012, 1900 to 2008, case-insensitive (all capitalizations summed), raw annual values. Each group is the sum of its terms, smoothed with a centered 7-year mean (±3 years), and indexed so the raw 1900-1909 mean is 100. The recomputed series match the supplied chart to within 0.05 index points for C1, C2, C3, C4, and C6. For C5 and C7 they differ by up to 5.9 and 3.0 points; both groups contain hyphenated terms (*self-control*, *self-sacrifice*, *self-denial*), which the Viewer reads as three tokens, and the supplied chart appears to have handled them differently. The brief uses the recomputed values.

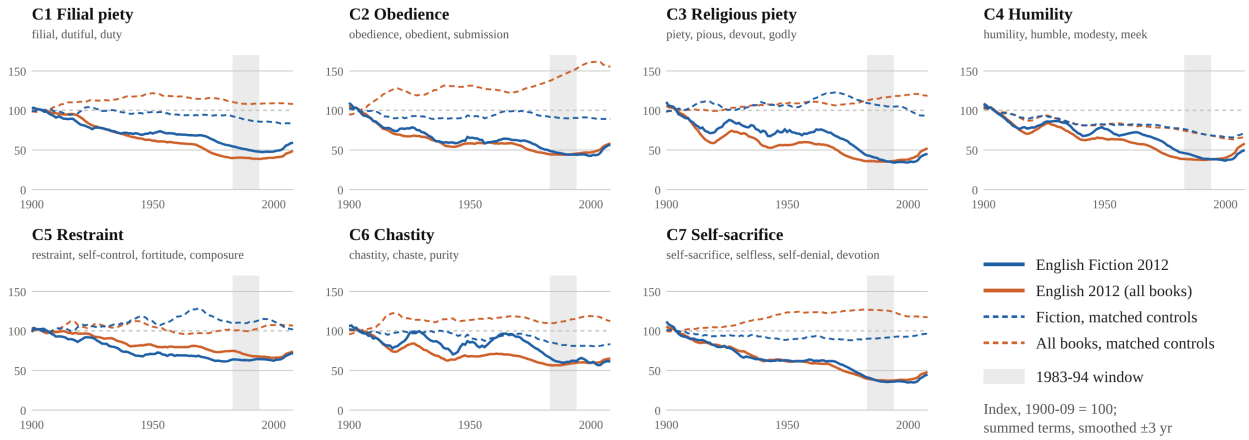

Figure 1. Summed frequency of each C1-C7 term group (solid) and of its frequency-matched ordinary-word controls (dashed), English Fiction 2012 (blue) and English 2012 all books (orange), 1900 to 2008, indexed to 1900-09 = 100, smoothed ±3 years. Grey band: 1983-94. Lexical frequency only, not framing. Source: Google Books Ngram Viewer, queried November 16 (groups) and November 18, 2026 (controls).

Table 1. Index values (summed terms; window means of the smoothed series, so 1900-09 reads 99 or 100 rather than exactly 100).

Group	English Fiction 2012				English 2012 (all books)			
	1900-09	1950-59	1983-94	2000-08	1900-09	1950-59	1983-94	2000-08
C1 Filial piety	99	72	51	53	100	61	40	44
C2 Obedience	99	62	46	49	99	59	45	52
C3 Religious piety	99	72	38	39	99	58	36	45
C4 Humility	99	72	42	42	100	63	38	49
C5 Restraint	100	70	64	67	100	80	71	69
C6 Chastity	100	83	62	60	100	69	58	62
C7 Self-sacrifice	100	62	37	39	99	61	38	43

Results.

- Every group fell in both corpora from 1900-09 to 1983-94: to between 37 and 64 in fiction, and between 36 and 71 in all books.
- From 1983-94 to 2000-08, all English books rose in six of seven groups (by 3.9 to 10.6 points; C5 fell 1.4). Fiction changed by between -2.3 and +3.1 points. The fiction lows fall between 1994 and 2003 for six groups (C5: 1980), with a rise in the last few years of the series.
- The supplied chart's note that fiction "keeps falling" fits the timing of the lows but overstates the 2000-08 window, where fiction is roughly flat.

Drift adjustment (new, November 18, 2026). Each term was matched to five ordinary words of similar 1900-09 frequency by the rule in section 2.3, here case-insensitive, 1900 to 2008, in the same two corpora (median gap 3%; largest a factor of 3.5). Each control was rescaled to its term's 1900-09 level, so every term keeps its weight. The control series is summed, smoothed, and indexed like the group. Adjusted index = group window mean / control window mean × 100. Intervals: 1,000 bootstrap draws of controls within terms.

Table 2. Drift-adjusted index (summed terms; 95% intervals for 1983-94).

Group	English Fiction 2012			English 2012 (all books)		
	Controls 1983-94	Adjusted 1983-94	Adjusted 2000-08	Controls 1983-94	Adjusted 1983-94	Adjusted 2000-08
C1 Filial piety	89	57 (49-68)	63	109	37 (25-72)	40
C2 Obedience	91	50 (41-62)	54	145	31 (24-42)	33
C3 Religious piety	107	36 (30-46)	41	116	31 (20-49)	37
C4 Humility	72	58 (42-83)	63	71	53 (40-77)	75
C5 Restraint	111	57 (46-72)	63	101	70 (55-95)	65
C6 Chastity	83	75 (63-96)	73	112	51 (37-90)	53
C7 Self-sacrifice	91	41 (30-59)	41	126	30 (16-62)	36

- The matched controls stand at 72 to 111 in fiction in 1983-94 and at 71 to 145 in all books, where five of seven are above 100; there the adjustment deepens most falls.
- After adjustment every group is below 100 in 1983-94 in both corpora, with every interval below 100: 36 to 75 in fiction, 30 to 70 in all books.
- The post-1990s contrast largely disappears. From 1983-94 to 2000-08, adjusted values change by -1.7 to +5.8 points in fiction (median +4.6) and -5.4 to +22.0 in all books (median +3.7). Only C4 still rises clearly more in all books (+22.0, against +4.8 in fiction).

Single-word dominance and equal weights. Table 3 gives each group's most common word and its share of the group's 1900-09 total, then two checks: an equal-weight index (each term indexed to its own 1900-09 mean, then a geometric mean; an arithmetic mean fails for C7 because *selfless* starts near zero), and a leave-one-out index without the most common word.

Table 3. Dominance, equal weights, and leave-one-out, 1983-94 (fiction / all books).

Group	Top word: share of 1900-09 total	Equal weights: raw	Equal weights: drift-adjusted	Without top word: summed, raw	Without top word: equal weights, raw	Without top word: summed, drift-adjusted (95% interval)
C1 Filial piety	*duty* 94.5% / 96.5%	53 / 45	65 / 44	55 / 47	54 / 48	70 (50-104) / 48 (37-67)
C2 Obedience	*obedience* 49% / 55%	48 / 44	53 / 37	54 / 54	54 / 49	56 (40-78) / 32 (23-44)
C3 Religious piety	*pious* 48% / 40%	40 / 38	38 / 35	41 / 38	42 / 41	32 (24-49) / 38 (28-54)
C4 Humility	*humble* 56% / 56%	46 / 42	60 / 63	51 / 47	51 / 46	64 (50-83) / 54 (37-91)
C5 Restraint	*restraint* 37% / 53%	59 / 57	52 / 53	67 / 53	60 / 50	49 (40-64) / 39 (28-59)
C6 Chastity	*purity* 64% / 75%	71 / 58	83 (70-104) / 52	81 / 59	83 / 58	95 (76-128) / 54 (32-98)
C7 Self-sacrifice	*devotion* 86% / 85%	73 / 81	56 / 32	54 / 56	94 / 107	53 (44-65) / 33 (23-53)

- **C1 and C7 are effectively single-word measures.** *Duty* is 94.5% (fiction) and 96.5% (all books) of C1's 1900-09 total, and *devotion* 86% and 85% of C7's, so their summed indices describe those words.
- **Where the fall holds without the top word and with equal weights,** raw and drift-adjusted: C2, C3, C4, and C5 in both corpora, and C1 and C6 in all books.
- **C1 in fiction:** without *duty*, *filial* and *dutiful* still fell (raw 55), but the drift-adjusted interval reaches 104, so the fall is not distinguishable from drift.
- **C6 in fiction:** without *purity*, the adjusted index is 95 (76 to 128); with equal weights, 83 (70 to 104). The fiction fall in C6 rests on *purity*.
- **C7:** the claim does not hold. Without *devotion*, the equal-weight raw index is 94 in fiction and 107 in all books, because *selfless*, near zero in 1900-09, rose about 8-fold (fiction) and 12-fold (all books), while *self-sacrifice* and *self-denial* each fell to about a third. Drift adjustment lowers these values (equal weights without *devotion*: 63 and 33), but only because the controls for *selfless* are words that were also rare in 1900 and rose later; that adjustment is fragile.
- The fiction-versus-all-books difference after the 1980s was noticed in the data, not specified in advance, and no significance test was run. It is reported as an observation.

2.5 Nineteenth-century British novels

Heuser and Le-Khac (2012) use 2,958 British novels, 1785 to 1900, "stretching far beyond the canon," with no popularity filter. Social restraint words fell "from ~0.30% of all word tokens ... to ~0.15%," which the pamphlet calls "a decrease of about 55%" (from a regression fit; the rounded endpoints shown imply 50%). The four "abstract values" fields together fell from about 1.0% of tokens in 1800-1810 to about 0.44% by the century's end, a 56% fall.

2.6 Character-level methods and bestseller lists

- **Underwood, Bamman and Lee (2018)** used BookNLP to identify characters and the words attached to them in about 104,000 works of fiction, mostly 1780 to 2007, mainly from HathiTrust, and checked the pattern on yearly *Publishers Weekly* bestsellers. Underwood's *Distant Horizons* (2019, ch. 4) extends this. They code gender, not virtue, but show that character-level series, including a bestseller subset, are feasible.
- **Archer and Jockers (2016)** trained a model on more than 20,000 recent novels to separate *New York Times* bestsellers from others, so popularity-labelled corpora can be built, at least for recent decades.
- **Hackett and Burke (1977)** compile annual US bestseller lists for 1895 to 1975, a sampling frame for dated popular samples.
- **Bestseller content analyses** found in the search compare best- with poor-selling novels (Harvey 1953) or study one 1998-2017 window (Gammaitoni 2021); neither abstract reports virtue shares at two or more dates.

2.7 COHA fiction section

Not run. English-corpora.org blocked scripted access (HTTP 403 challenge), and its free n-gram download needs registration and is not split by genre. No COHA figures are reported.

2.8 Pilot tests of the main test (new, November 17, 2026)

Two pilots ran the relative test of section 1 for C1-C7, with 10 works per cell. Rules, logs, texts, codes and scripts are in `litbrief-pilot/` (Pilot A: `results_pilotA.md`, `selection_log.md`; Pilot B: `pilotB/results_pilotB.md`, `pilotB/selection_log_B.md`; coding rules: `codebook.md`).

- **Samples.** Pilot A: for each year 1900-09 (Bookman lists, as compiled in Hackett and Burke 1977) and 1920-29 (*Publishers Weekly*), the highest-ranked bestseller with a Project Gutenberg text, each matched to a Gutenberg novel by an author never on an annual list 1900-39, from the same or previous year, with the same author sex, coarse genre, and US or British class, taken in a fixed seeded order. Pilot B: the same 1920s works (four matches replaced because they had no review entry) and 1990-99 *Publishers Weekly* bestsellers, matched to Wikipedia-listed American novels never on a PW annual list 1980-2009, by similar rules.
- **Measures.** Pilot A coded 715 extracts of about 70 words, sampled near cue words and principal characters, up to three per virtue per work. Pilot B coded one unit per work: the *Book Review Digest* entry (1920s) or the *Publishers Weekly* review (1990s; Kirkus for two works), capped at 450 words. Pilot B is the weaker measure: it codes what a summarizer reports, and its source differs by period.
- **Agreement.** A second pass by the same coder gave kappa 0.97 in both pilots (256 Pilot A extracts; all 40 Pilot B units). The same system coded both passes, so this shows consistency, not reliability.

	Pilot A: 1900s to 1920s, extracts	Pilot B: 1920s to 1990s, summaries
Mean share of works depicting each virtue: bestsellers; matched	0.31 to 0.29; 0.30 to 0.20	0.06 to 0.03; 0.13 to 0.01
Rarity difference in differences, pooled (95% interval)	+0.07 (-0.13 to 0.29)	+0.09 (-0.01 to 0.19)
Per-virtue rarity intervals excluding zero	C1 only (-0.5); not after correction for seven tests	None
Flaw-framed share of practising characters: bestsellers; matched	0 to 0.19; 0 to 0.07	0.50 to 0; 0.22 to 0
Flaw-framing difference in differences (95% interval)	+0.12 (-0.04 to 0.29)	-0.28 (-0.88 to 0.33)
Practising characters not evaluated	67 of 83 (81%)	2 of 17 (12%)

Reading. Both pilots are inconclusive for H1 and H1F: no difference was detected, and every interval includes zero, which at 10 works per cell is not a refutation. The pooled rarity differences have the opposite sign to H1 in both pilots. Flaw framing moves in the predicted direction in Pilot A and the opposite one in Pilot B, and every interval includes zero. Per-virtue intervals are about 0.5 to 1.1 wide, so with 10 works per cell only very large shifts could be detected. What the intervals do rule out: a pooled extra fall in bestsellers larger than 0.13 (Pilot A) or 0.014 (Pilot B) in the share of works depicting a virtue, and a pooled extra rise in the flaw-framed share larger than 0.29 (Pilot A; Pilot B's framing interval is unreliable). Per-virtue rarity intervals reach down to -0.86 (Pilot A) and -0.19 to -0.33 (Pilot B), so a large fall for any single virtue is not ruled out. The measures behave differently: 70-word extracts leave most framing unobserved, while summaries observe framing but record few virtues (1% to 13% of work-virtue pairs). On the 16 1920s novels coded in both pilots, extracts found a virtue in 26 of 112 work-virtue pairs and summaries in 9, with only 3 in common.

3. Interpretation (descriptive)

1. The relative test has been run only as two small pilots (section 2.8). Neither found a difference in differences for any virtue whose interval excludes zero after correction for seven tests, and the pooled estimates have the opposite sign to H1. This is inconclusive rather than a refutation: no difference was detected, but only large pooled effects are ruled out (section 2.8).
2. At term level, the decline in Kesebir and Kesebir's virtue words holds in fiction, where scientific texts matter much less, and is slightly more widespread there than in all books. Against frequency-matched ordinary words the median fall shrinks from 50% to 41% in fiction and from 46-47% to 33-37% in all books, and remains; for the 10 general terms it is not distinguishable from drift.
3. Term declines are not uniform: some moral sets rebound after about 1980 (Wheeler et al.), and the C1-C7 term groups rebound after the 1980s and early 1990s in all books more than in fiction (section 2.4), although after drift adjustment that contrast largely disappears, except in C4.

4. The methods run end to end on public texts and summaries (section 2.8); a coded popular corpus large enough to detect moderate shifts does not yet exist.

4. Interpretation (normative)

These are judgments, not findings, and Michael may weigh them differently.

- Fewer uses of *humility* or *sacrifice* in fiction is a fact about word use, not evidence that popular fiction frames those virtues as flaws.
- The next step that bears on the question itself: scale the pilot design to about 40 or more works per cell, with whole-chapter or longer readings so that framing is observed, and two independent human coders.

5. Hypotheses with confirm and refute conditions

#	Hypothesis	Would confirm	Would refute	Status
H1	Rarity: for at least one virtue, the share of popular works depicting it fell between two dates by more than in a matched corpus	Difference in differences below zero, interval excluding zero, same coding	Interval excluding zero with the opposite sign, or an adequately powered interval inside a pre-set equivalence margin (an interval that merely includes zero is inconclusive)	Inconclusive (tested in two pilots; intervals include zero; underpowered). No difference detected: pooled +0.07 (-0.13 to 0.29) and +0.09 (-0.01 to 0.19). Rules out a pooled extra fall larger than 0.13 (Pilot A) or 0.014 (Pilot B); per-virtue intervals allow falls up to 0.86 (A) and 0.33 (B)
H1F	Framing: for at least one virtue, the flaw-framed share among practising principal characters rose (or the endorsed share fell) between two dates by more than in a matched corpus	Difference in differences of the predicted sign, interval excluding zero, same coding	Interval excluding zero with the opposite sign, or an adequately powered interval inside a pre-set equivalence margin (an interval that merely includes zero is inconclusive)	Inconclusive (tested in two pilots; intervals include zero; underpowered). No difference detected: flaw-framing +0.12 (-0.04 to 0.29) and -0.28 (-0.88 to 0.33). Pilot A rules out a pooled extra rise above 0.29; Pilot B's interval is unreliable (2 to 9 characters per cell); no per-virtue framing estimate was possible
H2	Virtue and general moral terms declined in twentieth-century English-language books	Decline on the same list and corpus	No decline, or a rise	Supported for the 50 virtue words, raw and drift-adjusted (tested in Kesebir and Kesebir 2012; replicated here; adjusted median -37% in American English 2012 and -33% in English 2012, intervals exclude zero, section 2.3). For the 10 general terms the raw decline replicates, but the adjusted decline is not distinguishable from drift. Qualified by Wheeler et al. 2019: some sets rose after 1980
H2F	The same terms declined in the English Fiction subset	As H2, in fiction	As H2, in fiction	Supported for virtue words (tested in this brief): 41 of 50 fell, 1901-2000; drift-adjusted median -41% (-51% to -26%). General terms: 8 of 10 fell raw, but the adjusted median, -20% (-57% to +56%), is not distinguishable from drift
H2G	Term groups tied to C1-C7 declined from 1900-09 to 1983-94 in English Fiction 2012 and in English 2012	A lower index in 1983-94 for each group, in both corpora	No fall, or a rise, for a group	Supported (tested in this brief) for all seven groups in both corpora, summed and drift-adjusted (adjusted 1983-94: 36 to 75 in fiction, 30 to 70 in all books; intervals below 100). C1 and C7 are effectively single words (*duty*, *devotion*). With equal weights or without the top word: holds for C2 to C5 in both corpora and for C1 and C6 in all books; C1 and C6 in fiction not distinguishable from drift; C7 shows no raw fall without *devotion*. Lexical frequency only; does not test framing (H1F)

#	Hypothesis	Would confirm	Would refute	Status
H3	Moral-valuation and social-restraint language declined in nineteenth-century British novels	Decline in a defined novel corpus	No decline	Supported (tested in Heuser and Le-Khac 2012), as token frequencies
H4	A fall in virtue-term frequency is equivalent to a virtue becoming anomalous in popular literature	A shown correspondence between term frequency and coded portrayal in a popular corpus	A mismatch of constructs, or diverging measures	Argued against on the definitions in section 1; no empirical test
H5	Popular literature differs from a matched corpus in virtue-term use or portrayal in the same periods	A justified match, same measure, a difference	No difference, or the opposite	Untested as a planned comparison. The pilots (section 2.8) give within-period gaps of inconsistent sign across measures; the fiction-versus-all-books comparison (section 2.3) is not a popularity comparison

Candidates C1-C7 (Appendix A) specify H1F for named virtues. H2G tests only their word groups.

6. Verdict table

Item	Relative test (main)	Right unit and dates?	Verdict
Virtue portrayal in popular literature, rarity measure	Two pilots, bestsellers against matched novels, 1900s-1920s and 1920s-1990s	Yes, at 10 works per cell	Inconclusive at pilot scale: no difference detected; intervals include zero; underpowered
Virtue portrayal in popular literature, framing measure	Same two pilots	Yes, at 10 works per cell	Inconclusive at pilot scale: no difference detected; intervals include zero; too few practising characters
Virtue terms, all books, 20th century	Not a popularity match	Term level only	Decline measured, smaller but still present against matched ordinary words; not a test of the question
Virtue terms, English Fiction subset, 20th century	Fiction versus all books (genre, not popularity)	Term level only	Decline measured, smaller but still present against matched ordinary words; not a test of the question
C1-C7 term groups, English Fiction and all books, 1900-2008	Fiction versus all books (genre, not popularity)	Term level only	Decline measured in all seven groups, also drift-adjusted; C1 and C7 effectively single words; not a test of framing
Abstract-values fields, 19th-century British novels	Not run	Term level only	Decline measured; not a test of the question
Candidates C1-C7	Two pilots	Yes, small samples	Inconclusive: none confirmed, none refuted
The question as a whole	Pilots only	Yes, small samples	Open

7. Caveats, in one place

- **None of the term evidence is popular literature, and none is work- or character-level.** Google Books (fiction subset included) and the Heuser and Le-Khac corpus are library-like samples, and word frequency is not portrayal or endorsement. This caveat applies to every term result above.
- **General drift.** Ordinary words also drift in Google Books (Liberman 2018). Sections 2.3 and 2.4 now adjust against frequency-matched ordinary words chosen by a rule fixed before the data were fetched. Matching is on starting frequency and part of speech only, not on meaning, word length, or later trends. The filters missed 48 selected controls with moral, religious, or evaluative content; dropping them post hoc changed no adjusted virtue median. Controls for words near zero around 1900 (*selfless*, *selflessness*, *dependability*) are words that were also rare then and mostly rose later, so adjustments for such words are fragile. In all books the C1-C7 controls rose, so adjustment there deepens the falls. Technical senses and synonym shifts are not separated.
- **Comparability.** The rerun uses decade means for percent change; Kesebir and Kesebir used single-year endpoints read from graphs, so only the significance counts are directly comparable. English 2012 sets mix countries; no other Western language was examined.
- **C1-C7 term groups (section 2.4).** The term lists come from a supplied chart of unknown authorship and are not drawn from a published or rated list such as Kesebir and Kesebir's. C1 and C7 are effectively single-word measures (*duty* 94.5% to 96.5%, *devotion* 85% to 86% of the 1900-09 total), and the fiction fall in C6 rests on *purity*; the equal-weight and leave-one-out checks (Table 3) show which group results survive without the top word. An arithmetic mean of term indices fails for C7. Several terms have other senses (*duty* as tax or work shift, *submission* as a document, *devotion* in religious and romantic senses, *purity* in chemistry, *humble* and *composure* in everyday description). Hyphenated terms are read as three tokens. The 1983-94 window comes from the supplied chart's note, after the data were seen. Google Books sampling changed around 2000 (the reason Kesebir and Kesebir stop at 2000), and the post-1990s rise in all books falls partly in that period.
- **Sources read in part.** *The Bestseller Code* (catalog and reviews only), *Distant Horizons* (chapter list and summary), Harvey (1953) and Gammaitoni (2021) (abstracts only).
- **Search limits.** Semantic Scholar refused all queries; Google Scholar, JSTOR, and PhilPapers could not be searched by script. The November 12 searches were not logged, so every "not found" here rests on the November 16 searches (Appendix B).
- **Appendix A** comes from an unsourced assessment and is used only as hypotheses and a coding scheme.

- **Pilots (section 2.8).** Ten works per cell. One AI coder; the second pass was by the same system (in Pilot A on a 35% subsample chosen after the first pass), so agreement is inflated. No blinding to period, and famous bestsellers are recognizable. What counts as depicted depends on the extract cue lists (the C1 list was widened before coding, and that widening drives the one nominal C1 result) or on what summarizers chose to mention. Pilot B's source differs by period and is longer for 1920s bestsellers. "Less popular" means never on an annual top-ten list, and some matched authors were widely read (for example Grace Livingston Hill, Thomas Pynchon). Genre classes are coarse (Pilot B pairs *Hannibal*, categorized as horror, with a fantasy novel). Pilot B's matching and source-checking rules were tightened after the first matching run and before coding. Samples are the top eligible titles, not the top-ranked ones (for example, *Scarlett* had no checkable review). The 1895-99 lists were not checked for Pilot A's "never on a list" rule.

Appendix A. Proposed test design (unsourced; piloted in section 2.8)

Status. These seven candidates recast an unsourced assessment supplied for review, with no citations, corpus, dates, or counts; they support no verdict. Two small pilots (section 2.8) neither confirmed nor refuted any of them (inconclusive; underpowered). Each claims that a virtue is more often framed as a flaw in later than in earlier popular fiction, relative to a matched corpus, so each is tested with the framing measure (H1F), not the rarity measure. Section 2.4 tests only the frequency of related words.

Item	Earlier framing (hypothesized)	Later framing (hypothesized)
C1 Filial piety	Honoring parents as a duty	Entrapment; rebellion as growth
C2 Obedience to hierarchy	Respect for order	Conformist or complicit
C3 Religious piety	Purity and spiritual strength	Hypocrisy, fanaticism, or control
C4 Humility	Self-effacement as integrity	Self-assertion admired instead
C5 Emotional restraint	Endurance and self-control	Repression; disclosure needed
C6 Chastity	Honor and self-respect	Shame or arbitrary restriction
C7 Self-sacrifice	Highest form of love and duty	Codependency; low self-worth

How a test would run.

- *Sample:* novels from Hackett and Burke's annual lists at two or more dates; a matched sample of non-bestseller fiction from HathiTrust, same years and country.
- *Units:* each work (rarity measure) and each principal character who practises the virtue (framing measure).
- *Codes:* each practising character is coded *endorsed*, *flaw*, *mixed*, or *not evaluated* from the arc outcome, narrator evaluation, and other characters' responses, by two independent coders with reported agreement.
- *Decision rule:* confirmed only if the flaw-framed share rises (or the endorsed share falls) between dates by more than in the matched sample; no difference, or the reverse, refutes it.

Appendix B. Search log

All searches were run on November 16, 2026 (MT). Results were screened on title and abstract; a study was kept if it measured specified virtues or moral language in fiction or popular books at two or more dates, or supplied a method or sampling frame for doing so.

- **Set 1:** virtue bestsellers content analysis; moral values best-selling novels change over time; bestseller fiction values content analysis twentieth century; moral foundations fiction characters diachronic; character-level moral evaluation novels computational; Google Books English Fiction moral words; Corpus of Historical American English fiction morality virtue; Kesebir virtue words replication fiction; popular fiction self-sacrifice portrayal change; religious piety portrayal popular novels change; chastity sexual morality bestselling novels content analysis; humility pride portrayal popular fiction trend; emotional restraint stoicism fiction historical trend computational; moral language novels HathiTrust.
- **Set 2:** virtue bestselling novels; moral values best-selling fiction content analysis; character portrayal virtue popular fiction historical change; Google Books English Fiction moral virtue words; moral language fiction corpus diachronic.
- **Set 3:** Kesebir 50 virtue words list; content analysis values bestselling novels across decades; computational moral foundations fiction characters; COHA fiction virtue morality; each of the five sources the reviewer suggested.

Source	Time (MT)	Search terms	Outcome	Screened out, and why
Crossref API	5:50-5:55 PM	Set 1 (14 queries)	Top 10 per query reviewed. Kept: Harvey (1953)	Most hits were criticism of single authors, virtue philosophy, or other fields; none measured virtues across dates
Semantic Scholar API	5:50-5:57 PM	Set 1, then 3 shorter queries	HTTP 429 (rate limit) on every query	No results; not counted as searched
OpenAlex API	5:57 PM	Set 2 (5 queries)	Top 12 per query reviewed	Single works (Narnia, *The Hunger Games*), single genres or windows (teen romance; bestsellers 1945-1955), or unrelated corpora
General web search (in place of Google Scholar)	5:45-6:00 PM	Set 3	Found the full Kesebir text (author upload), Gammaitoni (2021), Jetter (2025), Ramezani et al. (2026), STONYBOOK (2023), a 1955-1959 bestseller thesis	Gammaitoni and the 1955-1959 thesis: one window, no virtue shares across dates in the abstracts. Jetter and Ramezani et al.: moral terms or word associations, not fiction portrayal. STONYBOOK: a resource for annotating novels, not a result
Google Scholar, JSTOR, PhilPapers	5:57 PM	virtue bestsellers content analysis	Scholar redirected scripted requests; JSTOR gave no usable results to a script; PhilPapers HTTP 403	Not searched

Source	Time (MT)	Search terms	Outcome	Screened out, and why
COHA (english-corpora.org)	5:48 PM	Site access	HTTP 403 browser challenge; n-gram download needs registration	Not run
Google Books Ngram Viewer	5:48-5:58 PM; 7:27 PM	Kesebir and Kesebir's 60 terms; the 26 C1-C7 group terms	Run (sections 2.3 and 2.4)	None
Crossref, PubMed, Open Library, library catalogs	5:45-6:00 PM	The five reviewer-suggested sources, plus Kesebir and Kesebir	All verified (Appendix C)	None
Pilot data sources (November 17, 2026)	12:30-1:50 PM	Wikipedia bestseller lists; Project Gutenberg catalog and texts (pglaf mirror, as gutenberg.org was down); *Book Review Digest* OCR (archive.org); PW and Kirkus review pages; Open Library ISBNs; Wikipedia year categories	Used for the pilots (section 2.8); every pick and skip is logged in `litbrief-pilot/`	Titles without a text or a validated review were skipped, as logged

Screened-out items cited by name above: Jetter, M. (2025), *The Economic Journal* 136(673), 351-369, doi:10.1093/ej/ueaf049; Ramezani, A., Stellar, J. E., Feinberg, M., & Xu, Y. (2026), *Nature Communications* 17, doi:10.1038/s41467-025-67891-2; STONYBOOK (2023), arXiv:2311.03614.

Appendix C. Sources and how each was checked

#	Source	How checked
1	Pechenick, E. A., Danforth, C. M., & Dodds, P. S. (2015). Characterizing the Google Books corpus: Strong limits to inferences of socio-cultural and linguistic evolution. *PLOS ONE*, 10(10), e0137041. doi:10.1371/journal.pone.0137041	Full text read
2	Kesebir, P., & Kesebir, S. (2012). The cultural salience of moral character and virtue declined in twentieth century America. *The Journal of Positive Psychology*, 7(6), 471-480. doi:10.1080/17439760.2012.715182	Full text read (author-uploaded copy); volume and issue checked in Crossref
3	Wheeler, M. A., McGrath, M. J., & Haslam, N. (2019). Twentieth century morality: The rise and fall of moral concepts from 1900 to 2007. *PLOS ONE*, 14(2), e0212267. doi:10.1371/journal.pone.0212267	Full text read
4	Heuser, R., & Le-Khac, L. (2012). A quantitative literary history of 2,958 nineteenth-century British novels: The semantic cohort method. *Stanford Literary Lab Pamphlet* 4. https://litlab.stanford.edu/LiteraryLabPamphlet4.pdf	Full text read; figures rechecked Nov 16
5	Greenfield, P. M. (2013). The changing psychology of culture from 1800 through 2000. *Psychological Science*, 24(9), 1722-1731. doi:10.1177/0956797613479387	Full text read (author lab copy); PubMed record

#	Source	How checked
6	Twenge, J. M., Campbell, W. K., & Gentile, B. (2012). Increases in individualistic words and phrases in American books, 1960–2008. *PLOS ONE*, 7(7), e40181. doi:10.1371/journal.pone.0040181	Full text read
7	Underwood, T., Bamman, D., & Lee, S. (2018). The transformation of gender in English-language fiction. *Journal of Cultural Analytics*, 3(2). doi:10.22148/16.019	Full text read (author copy); Crossref record
8	Underwood, T. (2019). *Distant Horizons: Digital Evidence and Literary Change*. University of Chicago Press. doi:10.7208/chicago/9780226612973.001.0001	Publisher record and chapter list; not read in full
9	Archer, J., & Jockers, M. L. (2016). *The Bestseller Code: Anatomy of the Blockbuster Novel*. St. Martin's Press. ISBN 9781250088277. No DOI	Catalog record and reviews; not read
10	Hackett, A. P., & Burke, J. H. (1977). *80 Years of Best Sellers, 1895-1975*. R. R. Bowker. ISBN 0835209083. No DOI	Library catalog records
11	Harvey, J. (1953). The content characteristics of best-selling novels. *Public Opinion Quarterly*, 17(1), 91-114. doi:10.1086/266441	Catalog record and abstract only
12	Gammaitoni, M. (2021). Building the collective imagination in the bestsellers of the Western world. *Intercultural Relations*, 4(2), 11-24. doi:10.12797/rm.02.2021.10.01	Abstract only
13	Liberman, M. (2018, October 9). Lexico-cultural decay? *Language Log*. https://languagelog.ldc.upenn.edu/nll/?p=40222	Read
14	Google Books Ngram Viewer, corpora en-fiction-2012, en-2012, en-US-2012, en-US-2009. https://books.google.com/ngrams	Queried Nov 16, 2026; control words Nov 18, 2026
15	"Virtue-term frequency: English Fiction vs all English books" (HTML chart supplied for review, Nov 16, 2026; author unknown)	Term lists taken from it; all values recomputed (section 2.4)
16	Wikipedia. *Bookman* list of bestselling novels in the United States in the 1900s (revision 1368937260); *Publishers Weekly* lists for the 1920s (revision 1380982495), 1980s (1363241603), 1990s (1369307667) and 2000s (1369088312). https://en.wikipedia.org	Wikitext fetched Nov 17, 2026; the 1900s page cites Hackett and Burke (1977), pp. 64-74. Used as pilot sampling frames
17	Project Gutenberg. Catalog (pg_catalog.csv) and plain-text editions of the 40 Pilot A novels. https://www.gutenberg.org (pglaf mirror)	Fetched Nov 17, 2026; ids logged in `litbrief-pilot/works.json`
18	*The Book Review Digest*, annual volumes 1919-1929. H. W. Wilson Company. OCR text at archive.org (item ids in `pilotB/selection_log_B.md`)	Fetched Nov 17, 2026; entries located by author and title and checked by reading
19	*Publishers Weekly* reviews (publishersweekly.com) and *Kirkus Reviews* (kirkusreviews.com), one page per 1990s pilot work (URLs in `pilotB/selection_log_B.md`)	Fetched Nov 17, 2026; each page checked for matching title, author and review date
20	Speer, R. (2022). *rspeer/wordfreq: v3.0* (v3.0.2) [Software]. Zenodo. doi:10.5281/zenodo.7199437	Package v3.1.1 used for the control-word pool (section 2.3)
21	Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), 216-225. doi:10.1609/icwsm.v8i1.14550	Lexicon (vaderSentiment 3.3.2) used to exclude evaluative words; DOI record checked

#	Source	How checked
22	Miller, G. A. (1995). WordNet: A lexical database for English. *Communications of the ACM*, 38(11), 39-41. doi:10.1145/219717.219748	WordNet 3.0 (NLTK) used for part of speech and proper names