

James F. English

Five Star Stories: *Readers and Ratings*

I have been studying the history of quantitative systems for rating works of art and literature. Prominent among these today are systems that express cultural value in stars and fractions of stars — more stars indicating a more positive judgment, a higher estimation of value. For shopping decisions in general, the dominant star rating platform is of course Amazon. But among platforms purpose-built for literature, the dominant player is Goodreads, Amazon's social book-reviewing subsidiary, which claims over 100 million members.

If prompted, many of those 100 million people would give the Goodreads star-rating system a rating of one star. Discussion threads both on the site itself and elsewhere in social book-chat media are rife with complaint and bewilderment about the curiously opaque and, on the face of it, unhelpful metric of “average user rating” for a book. Goodreads is scarcely unique in this respect: rating systems in general, and online rating aggregators in particular, have long been held suspect as devices for judging art and literature. Yet, despite the lack of trust placed in them, they have become the most ubiquitous cultural judgment devices of our era.

I'm not going to attempt a deep dive into the bowels of Goodreads in this essay, merely to offer a quick sketch of its place in the history of literary star ratings. What follows is a short story about ratings of short stories. The short story was the first form of literary work to which star ratings were systematically applied. This was in Edward J. O'Brien's *The Best Short Stories of 1915*, which inaugurated the annual *Best American Short Stories* series that continues to the present (Figure 1). The star rating system itself was pioneered a century earlier by the Englishwoman Mariana Starke in her *Letters from Italy* (1800; 1815) and *Travels on the Continent* (1820), as a concisely arithmetical way to present critical judgments of European painting and sculpture to middle-class British tourists. Starke's system involved exclamation

marks rather than stars (asterisks); but the latter became typographically standard as travel guides proliferated in the mid-1800s through the efforts of her British publisher, John Murray, and the German competitor, Baedeker. After its application to literature by O'Brien, the device was extended to film by Irene Thirer, who began to include a "star bar" in the header of her movie reviews for the *NY Daily News* at the dawn of the talkie era in 1928 (Figure 2).¹ A common view today is that star ratings are fit to evaluate ordinary consumer goods like office chairs or flashlights, but have no natural place in the domains of art and literature, where value is indeterminate or ineffable. But historically, aesthetic judgment provided the exclusive ground for the incubation and early adoption of these systems. It is only after they became a standard feature of judgment regimes across the major fields of artistic practice that the multi-tier rating or grading systems began to be applied beyond the arts: first, to other, less aesthetic kinds of "experience good" such as cuisine, for which Michelin launched its 3-star scale in 1931, and finally to ordinary goods and services like canned beans and cameras, which began in 1937 with the first annual *Buying Guide* from *Consumer Union*—the forerunner of *Consumer Reports*. The historical evidence suggests that an impulse to arithmetize the value of incommensurable and unmeasurable things—what Lucien Karpik calls *singularities*—is not imposed upon but is rather built into aesthetic ideology.²

Indeed, when O'Brien took up the star system from painting and the plastic arts and applied it (with manic enthusiasm and thoroughness) in literature, his aim was to advance an expressly anti-commercial, art-embracing agenda. O'Brien was part of the first generation of literary critics to center the short story—a quintessentially popular, ephemeral form—as the discipline's prime object of study, the exemplary form of literary art (and especially of *American* literary art). To resist what he saw as magazine editors' disabling constraints on the form, their encouragement of synthetic formulae and cheap plot hooks, O'Brien launched his annual review and anthology to steer readers toward the stories that were truly worth reading. The annual *Best Short Stories* anthologies O'Brien edited from 1915 until his death in 1941 included substantial "Yearbook" sections filled with lists and

1 The first review that included a star rating was Irene Thirer, "Port of Missing Girls' Film Gives Parents Moral Lesson," *New York Daily News*, July 31, 1928, 22. Thirer awarded the film one star out of what at that time was a maximum of three.

2 Lucien Karpik, *Valuing the Unique: The Economics of Singularities* (Princeton: Princeton University Press, 2010).

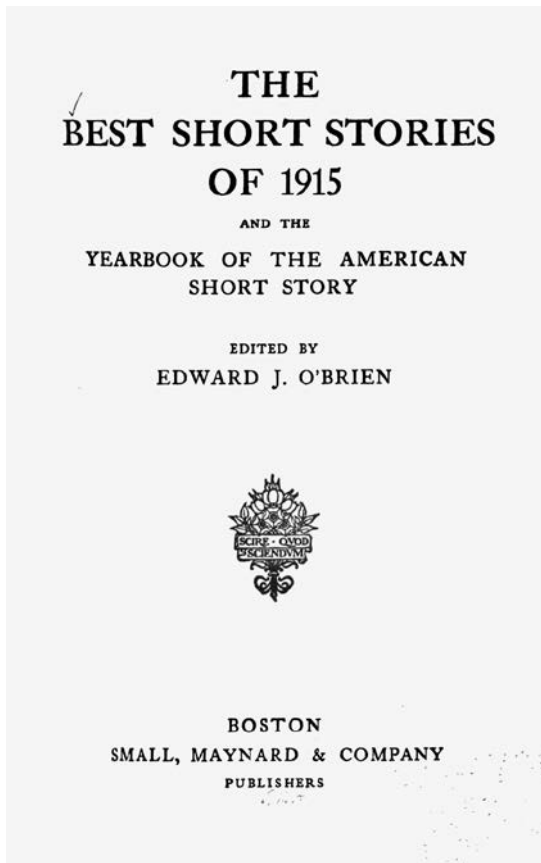


Figure 1: Title page from Edward J. O'Brien, ed., *The Best Short Stories of 1915*.

statistical tables in which he rated all the hundreds of stories published that year, awarding them zero to three stars based on a simple algorithmic syllogism of “substance” and “form”. Excellence in either of these aspects was worth one star; stories that excelled in both received two stars; and a third star was reserved for stories that successfully wove substance and form together in a unifying pattern of “spiritual sincerity”. These stories were listed in what O’Brien called a “special Roll of Honor”.³

3 *The Best Short Stories of 1915 and the Yearbook of the American Short Story* ed. Edward J. O’Brien (Boston: Small Maynard & Co., 1915), 7-8. O’Brien’s project of cultural renewal and its fate in the early years of the “Program Era”

INDEX OF SHORT STORIES

317

FINCH, LUCINE.

*The Boy at the Window. Harp. M. Oct., '14.

***The Woman Who Waited. Am. May, '15.

FINNEGAN, FRANK X.

And Your Neighbor As Yourself. Col. Jan. 16, '15.

Birdie's Mixed Motives. Col. Jan. 17, '14.

Birdie's Sane Christmas. Col. Dec. 19, '14.

Birdie the Strikebreaker. Col. March 28, '14.

From the Brink. Col. May 22, '15.

Marconi vs. Hymen. Col. Sept. 12, '14.

So Much Down. Col. Aug. 21, '15.

The Pulse of Spring. Lip. April, '15.

The Voice of Spring. Col. June 20, '14.

What the Siren Shrieked. E. W. July 26, '15.

FISHER, C. E.

The Riders. Col. Feb. 13, '15.

FISK, MAY ISABEL.

The Village Dressmaker. Pict. R. Feb., '15.

FITCH, ANITA.

***Colin McCabe: Renegade. Col. Jan. 9, '15.

FITCH, GEORGE.

Sam and I. Am. Nov., '15.

FITZGERALD, HENRY.

*The Pink 'Un. Col. June 5, '15.

FITZPATRICK, JAMES WILLIAM.

*The Hospital Ticket. Col. Dec. 19, '14.

**The Last Laugh. Col. Aug. 7, '15.

FLOWER, ELLIOTT.

Editor Parkin's Defeat. Pict. R. March, '15.

What Barnum Said. Cen. Dec., '15.

FLYNT, WENTWORTH.

The Mist. I. S. M. June 13, '15.

FOOTE, ELVERA.

The Escape of Cyrus. Sun. Oct., '15.

is well described by Kasia Boddy, "Edward J. O'Brien's Prize Stories of the 'National Soul'", *Critical Quarterly* 52.2 (2010): 14-28. Adrian Hunter's analysis of the critical debates around the short story in the early twentieth century suggests O'Brien's alignment with the "generalist" wing of literary criticism in its struggle against the "researchers" and their program of rigor and professionalisation. See Adrian Hunter, "The Short Story and the Professionalisation of English Studies" in *The Edinburgh Companion to the Short Story in English*, ed. Paul Delaney and Adrian Hunter (Edinburgh: Edinburgh University Press, 2018), 24-39.

THE ROLL OF HONOR FOR 1915

ALLEN, FREDERICK LEWIS.
Madame Zaranova.

ANONYMOUS.
Safety in Numbers.

ARCOS, RENÉ.
One Evening — The Meeting.

AUMONIER, STACY.
*The Friends.

BLACKWOOD, ALGERNON.
The Other Wing.

BROWN, ALICE.
The Return of Martha.

BROWN, KATHARINE HOLLAND.
The Old-Fashioned Gift.

BURT, MAXWELL STRUTHERS.
*The Water-Hole.

Figure 2: Page excerpts from O'Brien's alphabetized appendices: from the complete list of stories published in 1915, with star ratings (on opposite page); from the Roll of Honor for 1915 (on this page). Note special asterisk for the stories by Aumonier and Burt.

In explaining this putatively three-tier system, O'Brien took care to define zero stars as the first of *four* "natural [...] groups" (7), as well as including a distinctly higher fifth category consisting of three-star stories to which a special asterisk was added. The extra star marked them, he explained, as "so highly distinguished as to necessitate their ultimate preservation between book covers" (8), rather than merely in the ephemeral format of a magazine. His system thus actually consisted not of three ranks but of five—a number that seems to have exerted a certain gravitational pull on modern rating and grading regimes.

Looking at the star ratings in O'Brien's anthologies from the WWI years into the 1930s, one can be impressed by how well they track with the canon of twentieth-century American fiction as it was then taking shape. His Roll of Honor in the 1926 "Yearbook", for example, includes multiple stories by Sherwood Anderson, Kay Boyle, Willa Cather, Theodore Dreiser, F. Scott Fitzgerald, Ernest Hemingway, Zora Neale Hurston, Edith Wharton, and William Carlos Williams. As a credentialed expert offering judgments of a "generalist" bent to middle-class book buyers, O'Brien was not using the rating system in accordance with strictly personal values. While he claimed to per-

sonally read and rate every story published in an American magazine, he was of course well informed regarding the relative critical esteem of established authors as well as the reputational hierarchy of the magazines. Like Starke before him, he was a kind of human aggregator, condensing into an intuitive metrical scheme a complex of values shared by others in his wing of the expertise regime. Where he deviated from critical consensus, his tendency was progressive, as with the prestige parity he granted women authors, whose stories were, during his editorship, awarded stars and promoted to the honor roll in exactly equal proportion to men's.⁴ One looks in vain to find any other scale or scheme of literary value, prior to the present century, in which women authors were valued equally with men.

Through O'Brien's efforts and those of other advocates for the modernist short story, the core ambition of the *Best Short Stories* project was achieved: to sort short stories hierarchically, filtering out the ephemera and securing an echelon of timeless works bound for the library rather than the bottom of the birdcage. Over the course of the twentieth century, the short story became an increasingly important prestige form even as it lost commercial value and faded from the mass market. The "best" short stories offered a supply of modern "classics" for the training in criticism provided by postwar English studies, a "teaching canon." And as we know from the work of Mark McGurl, the short story came to serve also as the exemplary form for creative writing pedagogy in fiction workshops, the form *par excellence* of the Program Era.⁵

This is the point in our story about the rating of stories where Goodreads comes in. The star rating system is far more prominent on the literary field today than it ever was in O'Brien's time, but it has meanwhile become radically divorced from the scale of literary prestige and the program of the school. This is not because the millions of users on Goodreads are ignorant of the symbolic logic that grants short stories their place of special esteem. On the contrary. In the Price Lab at the University of Pennsylvania we've looked at the 1200 or so genres and subgenres Goodreads readers most frequently use to organize their book collections onto shelves: everything from "Anglo

4 I'm grateful to my research assistant Quinn Robinson for calculating the gender ratio of authors across the various levels of O'Brien's value system during his 25-year tenure as series editor. O'Brien maintained such a near-perfect balance between male and female authors that it is difficult not to assume a conscious social agenda. But he insisted his only criterion of excellence was unity of aesthetic and spiritual design.

5 Mark McGurl, *The Program Era: Postwar Fiction and the Rise of Creative Writing* (Cambridge, Mass.: Harvard UP, 2009).

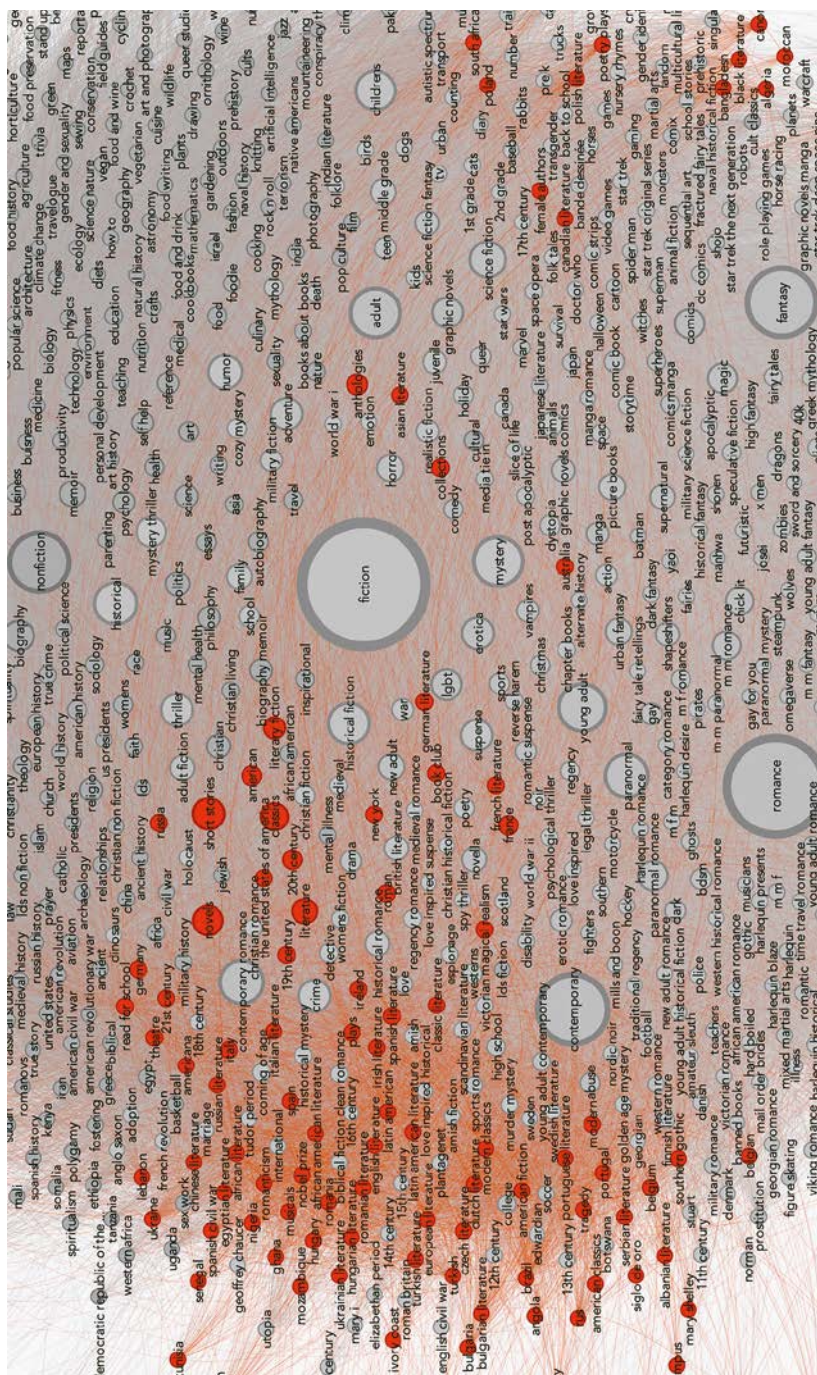


Figure 3: The network of 1200 common genre shelves in Goodreads, with shelves in the community of "literary fiction" highlighted in red. The size of a circle corresponds to how frequently it is used. (Visualization created in Gephi, by J. D. Porter.)

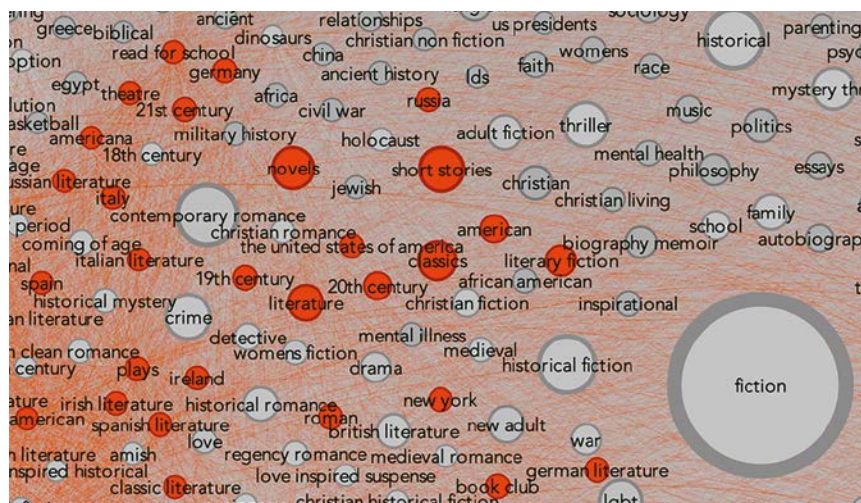


Figure 4: Zoomed in view of Figure 3, showing shelves in the community of literary fiction. “Short stories” is the largest, most frequently used shelf in this high-cultural genre neighborhood.

Saxon” to “zombies”. We built a network model where each of these genre-shelves is a node (a circle), sized according to its connectedness with other nodes, other shelves. The strength of connection between any two shelves depends on how often they co-occur in the shelving data of the same book. “Anglo Saxon” frequently co-occurs with “Medieval,” for example, but rarely if ever with “zombies”.⁶

By running what is called a community detection analysis over this network, we can discern algorithmically eight major genre-neighborhoods, tightly-connected node-clusters into which the millions of Goodreads users have placed their books. For convenience, we’ve given these major zones familiar genre labels: Fantasy & Science Fiction, Graphic, Historical, Literary, Mystery, Romance, YA, and Non-fiction. In the data visualization of Figure 3, created by J. D. Porter, all genre-shelf nodes have been left grey except those belonging to the community of “literary,” which are highlighted in red. As you see if

6 This data is based on the shelf-counts for a book’s ten most common genre-shelvings, as reported on Goodreads landing pages prior to the introduction of a new page format in 2022. The current site does provide access to complete shelf-counts, but to extract the top ten genre-shelves from that data would involve different methods than were used for the present paper.

Genre	Average Rating
Short Stories	3.79
Literature	3.76
School	3.74
Classics	3.86
Mystery	3.90
Historical Fiction	3.92
Romance	3.96
Fantasy	3.98

Figure 5: Average (mean) rating for books connected to eight different genre shelves in Goodreads. Ratings for genres associated with “literary” fiction are generally lower than for the major genres of popular fiction.

we zoom in (Figure 4), the largest node in this cluster, the subgenre most strongly interconnected to others in this community, is “Short Stories”.

Goodreads users tell us, through their collective shelving practices, that out of all the subgenres in the entire shelf array, it is “Short Stories” that they most strongly associate with the space of high critical esteem. At the same time, however, through their collective *rating* practices, they tell us that high critical esteem does not mean more stars. The average star rating of the 24,000 books shelved as “Short Stories” is 3.79 out of 5. That’s slightly higher than the average for the books shelved as Literature (3.76) or School (3.74), and slightly lower than the average for books shelved as “Classics” (3.86). But this entire genre neighborhood, the zone of canonicity and critical prestige, is rated lower than all the other major neighborhoods. The average rating of books connected to “Mystery” is 3.90, “Historical Fiction” 3.92, “Romance” 3.96, and “Fantasy” 3.98 (Figure 5).

It is also the case that, among short story books, the “best short stories” are not rated higher than average. The average rating of the most recent ten volumes of *Best American Short Stories* is 3.79—which makes them, on this metric, no better than average for books connected to the Short Story shelf. In fact, I have found that in general books that win critical recognition as “best” in any given genre (e.g.

books shortlisted for mystery novel prizes like the Edgar or science fiction prizes like the Hugo and the Nebula) tend to be rated *lower* on Goodreads than the average book in that genre: 3.83 for prizelisted detective novels vs. 3.93 for non-prizelisted; 3.82 for prizelisted science fiction novels vs. 3.93 for non-prizelisted.⁷ Even books that stand out in a given genre as *bestsellers*, best by the measure of commercial value, tend slightly to trail the average rating.⁸

In short, between a book's Goodreads rating and its position in the most relevant hierarchies of value—its canonicity (value in the academic system), its mainstream prestige (value conferred by prizes and awards), or its popular success (commercial value, number of ratings in Goodreads) there exist more inverse correlations than positive ones.

Aggregation—the crowd-sourcing of judgments—cannot in itself account for the misalignments between Goodreads' star ratings and other judgment devices of the literary expertise regime. Why the sky-high ratings for poetry compared to YA romance? Why is *Pride and Prejudice* rated so much higher than *Romeo and Juliet*, *Anna Karenina* so much higher than *The Great Gatsby*? One key to understanding the shift from the original star rating systems like O'Brien's to ratings aggregators like Goodreads is the elimination of the zero-star option. For O'Brien, like Starke, zero stars was the norm, covering the whole range of cultural value from appallingly bad to well above average. One star was already an exclusive attainment, and three stars was reserved for works of rare quality. Of the 2,200 stories O'Brien rated in 1915, only 93 (or 4%) were awarded three stars and placed on the Roll of Honor. About half of those (2% of all published stories) appear in the appendix with a special asterisk, a fourth star, denoting extra high distinction.⁹ And only about half of *those*—20 stories (less than 1%)—were finally selected for reprinting in the anthology. "Honor roll" is indeed an apt term for works thus distinguished. The

7 Based on a 2018 analysis with Scott Enderle at the Price Lab of winning and shortlisted novels for leading prizes in those two popular genres, compared to samples of 100 other novels in each genre.

8 My 2023 analysis, with J. D. Porter, of more than 600,000 books in Goodreads found a slight positive correlation between the number of ratings of a book (its popularity) and its average star rating. But this does not contradict my earlier finding in the Contemporary Fiction Database Project, that the very top best-sellers for each year dating back to 1960 tend to have lower average ratings than other novels in Goodreads. That study also found that novels shortlisted for major novel-of-the-year awards had even lower average ratings than the best-sellers.

9 This fourth level of the system, the three-star-plus-extra-star level, was discontinued in 1922 without, so far as I know, any statement or rationale from O'Brien.

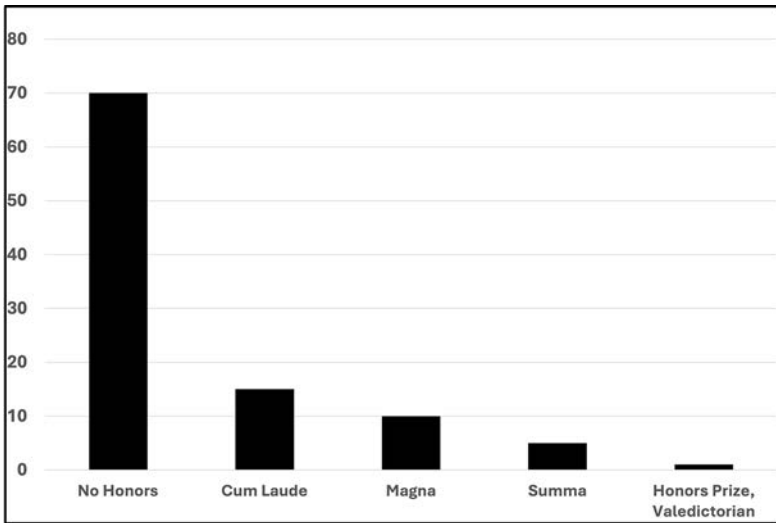
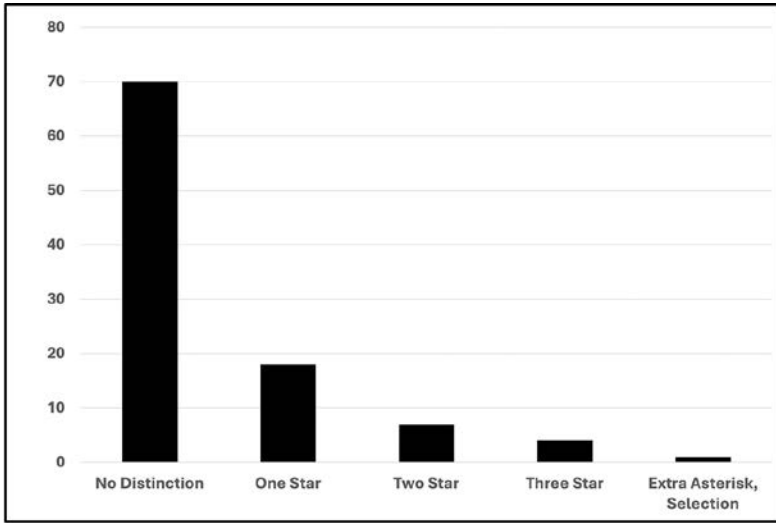


Figure 6: Percent distribution of stars, none to four, in O'Brien's 1915 volume (above); percent distribution of academic honors, none to valedictorian, in the US university system, 2015 (below).

original star rating systems functioned as *scales of exceptionality*, homologues of the Latin Honors system in higher education, with a sharply declining fraction of recipients at each higher level of honors (Figure 7).¹⁰

Contrast this with Goodreads. Deprived of the zero-star baseline, Goodreads users have to make room in their five-tier distribution for all those run-of-the-mill, “not bad” or “ok” books — the vast majority — as well as the ones they judge “terrible” or “unreadable,” which are now assigned, as stigma, the one-star rating that originally signified esteem. The result is a distribution resembling not Latin honors but letter grades in the age of grade inflation: a rising curve on which the vast majority of values are either A or B, 5 or 4 (Figure 7).

The Goodreads rating system is a scale of *negative* exceptionality. Though superficially resembling the systems of O’Brien and other pioneers of *cultural* rating systems, it in fact derives more closely from the rating schemes developed decades later by *Consumer Reports*. For users of CR’s *Buying Guides*, it was more essential to distinguish items found to be “poor” or “substandard” (Consumer Reports’ two lowest categories) than finely to differentiate among the highest-end luxury goods. As a review aggregator, Goodreads operates on quite different principles than *Consumer Reports*, but its rating system makes this decisive accommodation of negativity. It provides reviewers with a sharper tool for indexing their disappointment than their esteem.

Disappointment explains, in part, why more prominent books (prizewinners, bestsellers, school texts, classics) tend to score lower than average on Goodreads. The visibility and symbolic elevation that these books have attained through other judgment devices (whether academic or commercial), attract readers who would not normally be reading books in that neighborhood, or on that particular shelf: readers who are more likely to be disappointed. And given the compression of scores toward the top of the scale (nearly $\frac{3}{4}$ of all ratings in Goodreads are 4’s or 5’s), disappointed readers enjoy disproportionate power. A one- or two-star review lowers an overall rating more than a five-star review can raise it. Again, academic grades provide a familiar analogy. A single F on a transcript drops a student’s GPA more than an A can boost it.

¹⁰ O’Brien was of course intimately familiar with the Latin Honors system, which was first introduced at his alma mater, Harvard, in 1869. How consciously his rating system was modeled on Latin Honors rather than, for example, on the star ratings in Baedeker guides, I am unable to say.

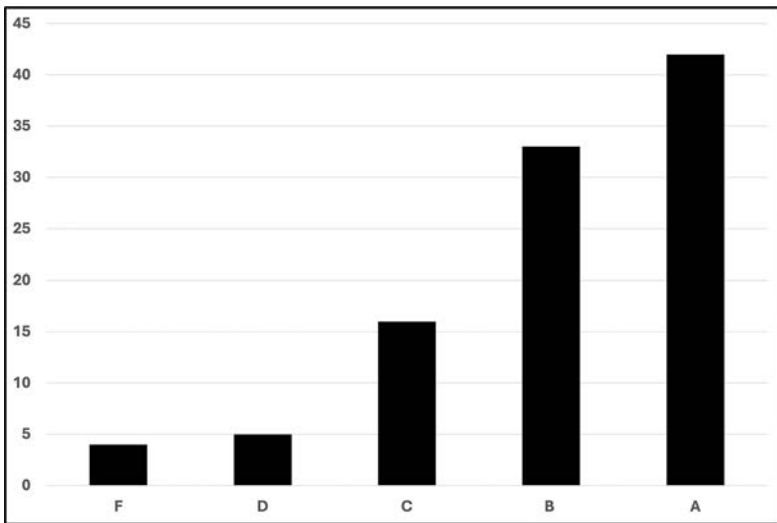
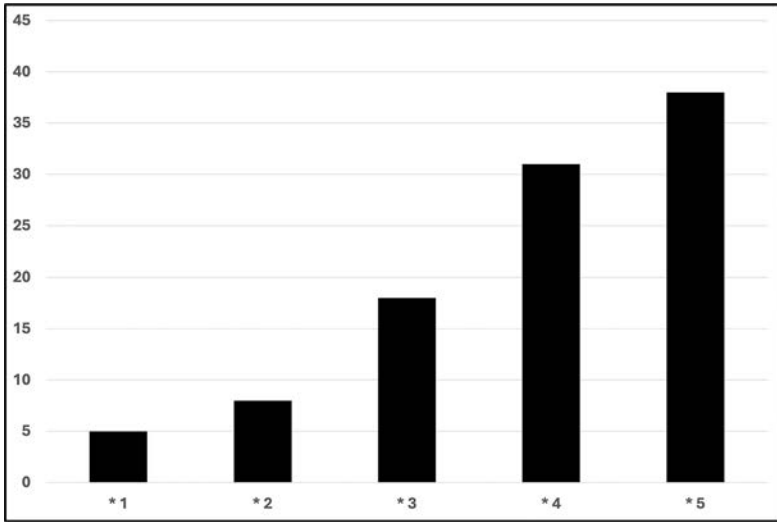


Figure 7: Percent distribution of stars, one to five, in Goodreads (above); percent distribution of academic grades, F to A, in U.S. higher education (below). Goodreads data based on 1800 novels in the Contemporary Fiction Database Project at the Price Lab, University of Pennsylvania. Source for academic grades: Catherine Rampell, "A History of Grade Inflation," *New York Times*, July 14, 2011.

This doesn't mean readers can't use Goodreads' star ratings and accompanying distribution-chart graphics to help guide them toward a book they'll love. There are well-honed strategies for doing that. But these strategies generally entail more scrutiny and assessment of the one-star reviews than the five-star, further amplifying the influence of negative judgments within the site, elevating them in Goodreads' second-order hierarchies of "top" reviews and "top" reviewers. Effective navigation strategies also lead users away from the system of star ratings into other features and affordances of the site such as ranked lists and curated sets of favorites. The arithmetical ratings themselves are simply not aimed any more at capturing "the best", but rather, by activating the core negative constituents of taste — aversion and avoidance — at keeping readers happily within the bounds of their established preferences.