The Entropy of Television

Katie Lin (celin)
May 4, 2023

It is difficult to overestimate the outsized impact that television has had on contemporary American society. And while it is undeniably influential, the history of television has been fraught with both critical and cultural antipathy. Its structure and form have been, and continue to be, undergoing constant evolution with the advent of platforms like YouTube and the emergence of the "streaming wars", the rise of limited series and experimental narrative forms, and the ever-shifting cultural dynamic with content consumption and its harrowed relationship with human attention spans.

At least until the turn of the century, when it began to undergo a comprehensive cultural re-evaluation, television was viewed as "a disposable product, like a Dixie Cup," by the general public.[1] However, the academic study of television developed during the 1970s and 80s, with scholars approaching the subject from both a textual analysis lens as well as a sociological perspective.[2] The social aspects of the medium, such as its educational advantages or its negative psychological effects, are of great interest to media theorists, but continue to be the subject of debate and further research.[3]

Like television, the Flynn effect is a topic that is both fascinating and perplexing to researchers and scientists alike. The Flynn effect is a well-documented biological and sociocultural phenomenon that refers to a continuous and approximately linear increase in intelligence quotient (IQ) test scores in the 20th and 21st centuries. It has been estimated that scores increase by about 3 points per decade, although the tests themselves are designed as specific measures of intelligence relating to "abstract problem-solving" abilities.[4] The Flynn effect is subject to numerous ongoing debates over possible explanations for the trend in rising scores, while others hypothesize that the sustained increases may have plateaued and IQ scores are now beginning to fall, illustrating a reverse Flynn effect.

While these are seemingly disparate topics, the rise in intelligence surprisingly shares some key commonalities with the development of television in the United States. Both concepts emerged during the 20th century and began to take form around the 1950s due to a growing number of television stations and the publication of the Wechsler Adult Intelligence Scale (WAIS), the most widely used IQ test worldwide.[5][6] As media and technology increasingly affect and transform the way we relate socially, economically, politically, and epistemologically, our relationship with television becomes a convincingly worthwhile investigation.

Specifically, the motivation of this paper is to answer the following question:

---

1 https://www.npr.org/2019/06/25/735514087/i-like-to-watch-is-a-passionate-brilliant-defense-of-tv
2 https://en.wikipedia.org/wiki/Television_studies
3 https://en.wikipedia.org/wiki/Social_aspects_of_television
4 https://doi.org/10.1016/B978-0-12-815744-2.00008-2
5 https://stephens.hosting.nyu.edu/History%20of%20Television%20page.html
6 https://en.wikipedia.org/wiki/Wechsler_Adult_Intelligence_Scale

Can we track the Flynn effect through the evolution of television?

Of course, it is highly improbable that the rise in IQ scores is directly linked to increasing complexity in television plots and narratives; however, a relationship between the two concepts may signal a cultural shift in intelligence and complexity as exemplified by the TV shows we watch.

The majority of this investigation is rooted in Information theory, the science of signals. In order to quantitatively investigate this medium, I will primarily be using Claude Shannon's formula for calculating information entropy, which he introduced in his 1948 paper, "A Mathematical Theory of Communication"; this work is largely recognized as the foundation of information theory as a field of study. Sometimes referred to as Shannon entropy, this calculation is a measurement of the average level of "information", "surprise", or "uncertainty" for a variable (in this case, an episode of television) as computed through the use of probability distributions.

One particularly interesting application is the entropy calculation of English language novels, which acts as a proxy for the information density of a text. Jane Austen's *Pride and Prejudice* (1813), a pre-Victorian era novel, has an entropy of 9.06 bits. Virginia Woolf's *To The Lighthouse* (1927), a modernist work, has a comparative entropy of 9.13 bits.[7] This relative consistency in information density is nearly constant for all English language novels, regardless of publication date, era, style, or content, which lends itself to the conclusion that the richness of our vocabulary, at least for the English language, is generally unchanging.

For my data collection process, I took inspiration from a variety of projects and methodologies.[8][9] I aimed to strike a satisfactory balance between maximizing the utility of the data and optimizing the breadth of artifacts. For my selection of television shows to analyze, I began with this Wikipedia page of the top-rated United States television programs by television season. I took note of ones that recurrently appeared for at least a few seasons, using ratings as a proxy for cultural relevance. As I neared more recent seasons of television, it became apparent that these ratings alone would not be sufficient for a number of reasons. While variety shows were some of the first to successfully crack the formula for American television broadcasting, their stronghold over the television circuit quickly became diluted by the introduction of new genres and narrative forms. As a general rule of thumb, I decided to disqualify all variety shows and anthology series due to their lack of continued narrative arcs and consistent casts of characters. Additionally, after reviewing several lists of popular and influential television series, I decided to narrow my focus to the comedy genre (including its myriad subgenres and hybrid types) as my primary subject of analysis.

Using a list of 62 shows spanning the 1950s to the present, I whittled down my final selection to 43 shows based 1) on the availability of transcripts from the site I used to obtain episode transcripts: https://www.springfieldspringfield.co.uk/, and 2) on the distribution of shows by decade, which initially favored the inclusion of more contemporary series. I did not consider any currently running shows, although I think they could have provided another interesting layer to

[7] Simon's lecture on Information Theory. Feb 20, 2023.
[8] https://github.com/luonglearnstocode/Seinfeld-text-corpus
[9] https://github.com/BirkoRuzicka/Star-Trek-Transcripts

my data analysis, and the dataset I collected does not include all episodes of every show. For shows that aired for more than 6 seasons, I limited my data collection and analysis to the first 6 seasons. As time was a significant constraint, I standardized and simplified the respective years that a show aired for to the decade in which the show first began to air.

Of the final 43 shows selected, the distribution of data with respect to decade is as follows:

	# of series	# of episodes
1950	1	127
1960	6	604
1970	6	369
1980	6	558
1990	8	1096
2000	8	841
2010	8	513
	43	**4108**

In total, I gathered 4,108 episodes of television to analyze, which gives us an average of 95.534 episodes per show. Of course, not every series I scraped transcripts for had 100 episodes or even close to that many, but this was a helpful indicator that I had a fairly large enough dataset to be able to calculate and use statistical averages, which partially helped compensate for individual episode transcript errors or outlier values.

I wrote all of my code for processing, scraping, cleaning, and analyzing my data in Python 3, using a combination of Jupyter notebooks and individual .py files. In addition, I also leveraged the power of a number of Python packages, including pandas, BeautifulSoup, and re (regex).

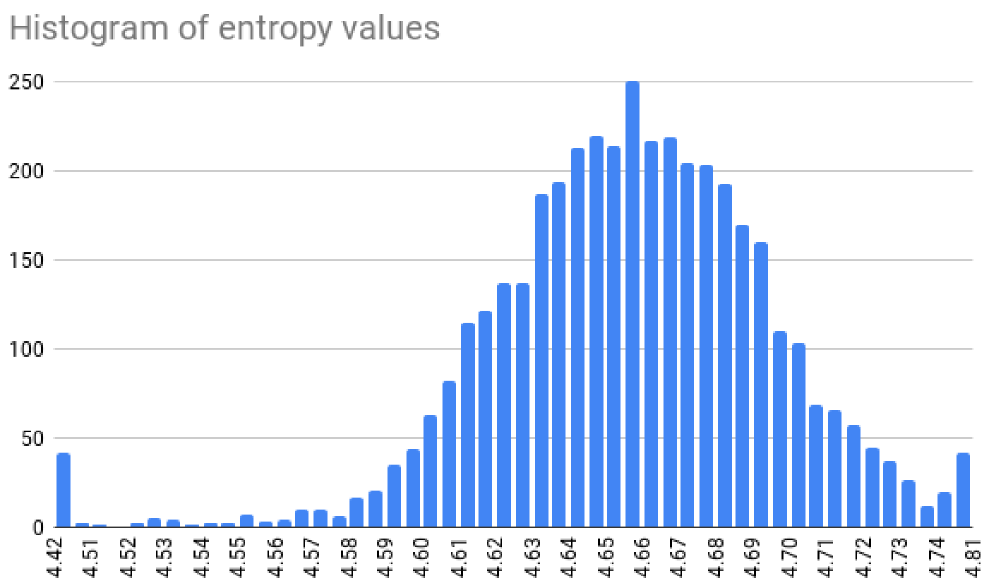
I organized my data by decade, show, and season. Every season of a show had its own .txt file inside of its respective decade folder. Episodes were delimited in the .txt file by unique episode markers, which I used to separate them from each other when processing a season's worth of transcripts.[10] Because my dataset includes over 4,000 episodes, there are a number of spelling and formatting errors. However, the majority of the transcripts are readable and accurate enough to get close estimates of actual entropy.

Looking at just the entropy of all 4,108 individual episodes, I first calculated the average, median, min, and max values:

[10] Here, I define transcripts as the audible dialogue and speech of an episode of television. Lines of dialogue were not assigned by character, and any parenthetical information or additional metadata was removed using custom text cleaning scripts.

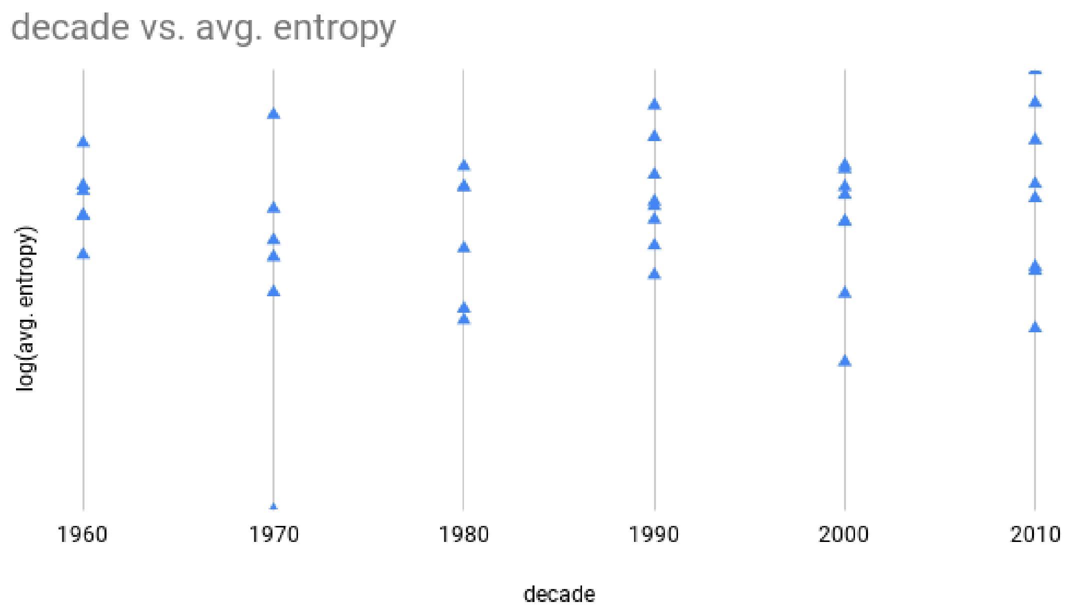
Average	4.657565702
Median	4.658852267
Minimum value	4.422149859
Maximum value	4.810206724

The entropy follows a normal curve with a slight negative skew and has outliers of both ends of the curve.



In order to answer my research question, I grouped the data by decade and calculated the average entropy for each show. When looking at this data, I found that by omitting the 1950s, which only had data for one show, some more interesting trends emerged.

The below figure is a scatter chart of the average entropy on a log scale with respect to the decade the show first began airing during. As time passes, there seems to be a larger range of entropy values for television shows. This increasing variance may be a signal of a widening cultural gap that is being reflected through the media we consume.



Overall, the average entropy stays largely unchanged over time, much like the English language novels mentioned earlier. This does not rule out the possibility of finding significant differences in the information density of television shows over time, but it does indicate the potential need for a different statistical calculation in order to identify a signal.

When processing Austen's *Pride and Prejudice* and Woolf's *To The Lighthouse*, the interesting insight that emerges is the significant difference between their mutual information. Mutual information, also known as relative entropy or Kullback-Leibler divergence, is a measure of shared information content. So, given what we know about the probability distributions of individual words in a novel, mutual information measures how much one word can tell us about another one.[11] The mutual information of *Pride and Prejudice* is about 1 bit greater than *To The Lighthouse*; this means that it is easier to predict the next word of *Pride and Prejudice* when given the preceding word. Woolf's text has more uncertainty because it sends fewer signals indicating what might come next.

I thought that I might be able to find a clearer pattern or signal when looking at how the information in an episode relates to itself. For my investigation, I used conditional entropy to analyze the data. Conditional entropy is what is used to calculate the mutual information of a variable; the mutual information between some variables X and Y is the entropy calculation of Y minus the conditional entropy of Y given that we know variable X. A higher conditional entropy leads to a lower mutual information, which indicates a lower ability to predict what comes next.

As a quick aside: my conditional entropy calculations may look a bit odd at first glance. My implementation for calculating the conditional entropy of an episode looked at each individual character in an episode's transcript, rather than each word. If I had a bit more time, I would've rewrote my calculator to split transcripts by words, however, I think my data and calculations still get at the larger idea of decomposing an episode into its information and measuring how the individual pieces relate to each other.

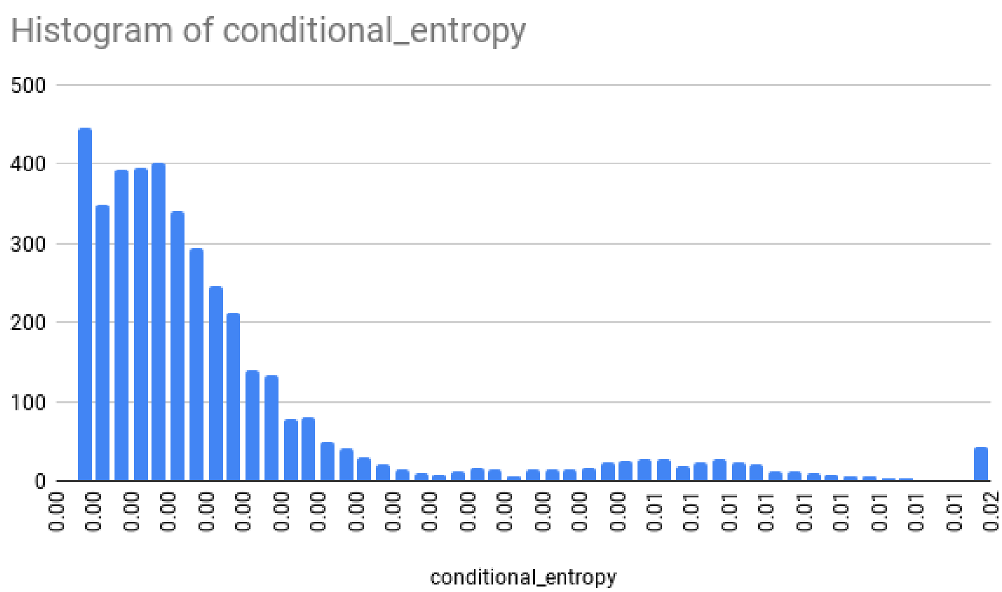
[11] https://en.wikipedia.org/wiki/Kullback%E2%80%93Leibler_divergence

After getting each episode’s conditional entropy, I calculated the average, median, and max values:

Average	0.00138004721051986
Median	0.00085127676565127
Maximum value	0.02008338828546140

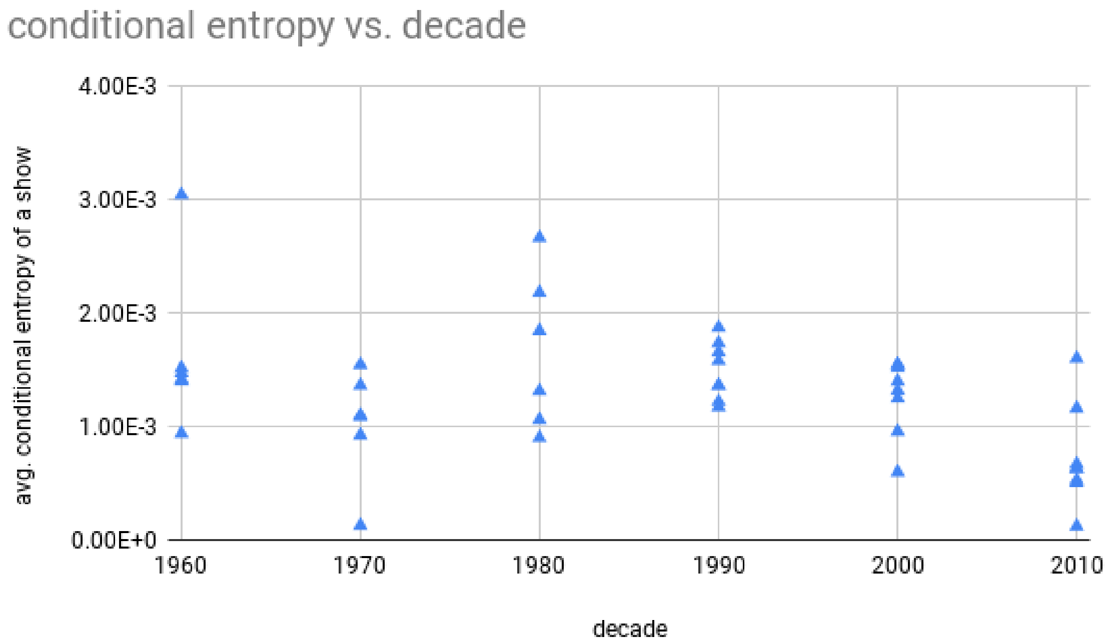
In contrast to the traditional entropy summary statistics, here, the mean conditional entropy is greater than the median, and they have a larger relative difference than the traditional entropy mean and median.

For these calculations, there was a much higher likelihood and occurrence of error due to transcript misspellings and formatting issues, which directly impacts the conditional entropy. Because of this, I did not include the minimum value as part of my summary statistics because the minimum value in my dataset was zero. I don’t know for sure if that is a true calculation of the conditional entropy, although it is fairly unlikely that is the case. Regardless, I chose to focus on the statistics I could gather.

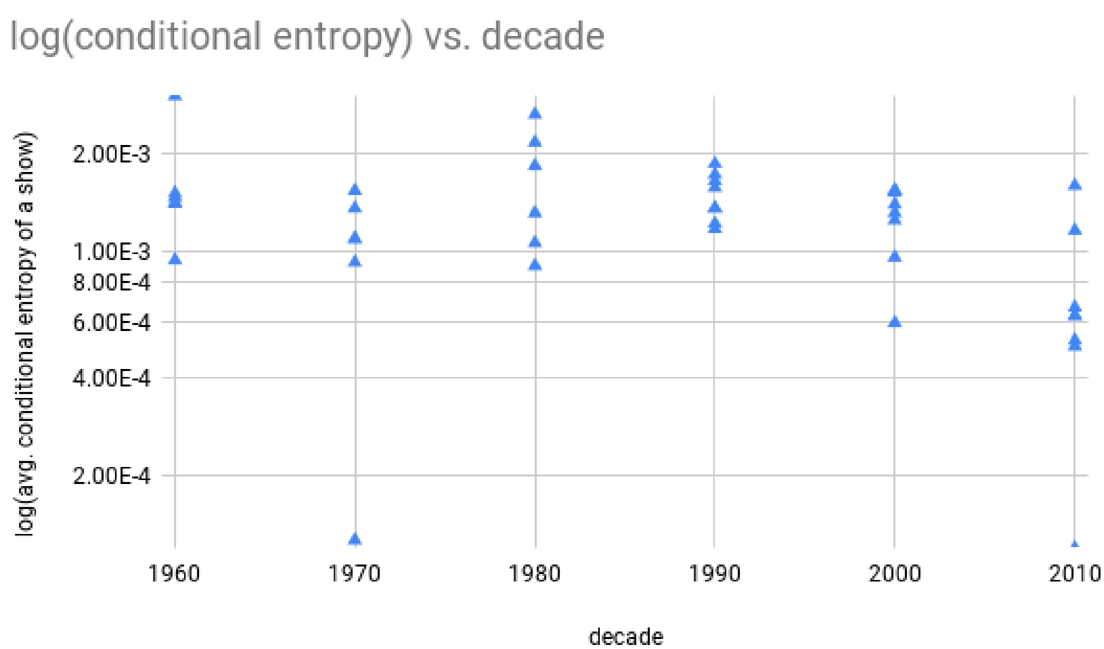


Again, comparing the conditional entropy distribution to the traditional entropy distribution, we can see that there is more variability in the conditional entropy of an episode, which aligns with the English language novel findings.

To address my original research question, I then grouped the data with conditional entropies by television show, calculated the average conditional entropy, and plotted it with respect to time. Like before, I’ve chosen to omit the 1950s in order to better identify any signals or trends.



The above figure is a scatter plot of the average conditional entropy of a show vs. the decade the show first began airing during. The below figure shows the same data, but with the average conditional entropy plotted on a log scale to identify significant deviations and data points.



Both figures indicate a general increase in conditional entropy until 1980, where it peaks, then an overall decline in conditional entropy from 1980 to the present. Moreover, there seems to be a significant level of variance in the conditional entropy range for a given decade. The one outlier in 1960 could potentially be a result of transcript errors, which would make for a much clearer signal overall in the data. If we ignore that one maximum value in the 1960s, the conditional entropies for all other shows are fairly similar. When we advance by a decade, the conditional entropies begin to separate and become more distinct. In the 1980s, where we see an overall peak in conditional entropy, we also see the greatest amount of variance and the largest range of values.

The 2010 data is interesting to compare and contrast with 1980; while 2010 overall has lower conditional entropies than 1980, the range and spread for 2010 shows is similar to that of 1980 shows. When going back to the traditional entropy calculations and its scatter chart, we see a somewhat familiar pattern in our conditional entropy spreads, but it is clear that the conditional entropy is a more revealing and indicative measure of information density and uncertainty than entropy alone.

In summary, there is compelling evidence for the existence of the Flynn effect, coupled with the reverse Flynn effect, as seen through the rise and peak in the conditional entropy of television shows alongside the more recent increase in deviation between television shows from the same decade. As I hypothesized earlier, the increasing variance in entropy, as shown in the histogram of entropy distributions for all episodes, hints at a widening cultural gap. My findings from my conditional entropy calculations, particularly the scatter chart that uses a log scale for conditional entropy, further support this hunch; earlier decades have more similar conditional entropies, while more recent decades indicate an expanding range of offerings.

Additionally, if we reexamine the log of conditional entropy figure with respect to the timeline of the Flynn and the reverse Flynn effect, there is evidence that an overall high conditional entropy is correlated with increased intelligence capabilities, while more variance in conditional entropy indicates a lower overall average in intelligence capabilities, which aligns with observations of the reverse Flynn effect.

Since the history of television is relatively short within the grand scheme of humanity, it’s hard to extrapolate cyclical patterns or trends that may be a part of the evolution of the medium with respect to culture. However, further explorations of this idea using existing data could look more closely at the data with respect to the time of an episode’s production and release. Other investigations could examine the structure and form of television shows and episodes in addition to the textual analysis. For instance, the popular “mockumentary” style sitcom may have significant differences in overall and conditional entropy due to their heavy reliance on facial expressions, talking heads, and cutaway gags with visual comedic effects. All in all, the subject presents a number of interesting questions that reflect the state of culture and society at large.