

## To Finity, Not Beyond

*How many stories are there?*

I heard a writer say on the radio the other day that there are “an infinite number of stories we can tell.” That’s a comforting thought for a screenwriter. But it isn’t true.

A 120-page script pretty well covers all TV and movie stories. With a generous 60 spaces to the line and 48 lines to the page, that’s a maximum of 345,600 characters.

A script, though annoying without capitals, is still legible in lower-case. Many of the ones I read get by without punctuation, but let’s assume nine additional marks beyond our 26 letters, including the space.

Any script we write must have one of those 35 characters in the first position on page one, one of those 35 characters in the second position, and so on.

The total number of scripts with a length of one character is 35. The number of two-character scripts is  $35^2$ , or 1,225.

So the number of possible 120-page screenplays is  $35^{345,600}$ , a number that can fairly be described as staggeringly large, but still finite. It’s smaller than many numbers commonly used by mathematicians.

Call our big number **S**, for Stories. And consider all of the scripts with fewer than 120 pages – sitcoms, animation – to be the special cases in which the characters after a certain page are all spaces.

Those **S** scripts include the story of my sitting down and writing this article, including every conceivable way I could have written it, and with every English word taking the place of every word you’re now reading, in every imaginable (and, I suppose, unimaginable) combination.

They include the story of you – no matter what your nature or circumstances – sitting down and reading it. Or standing up, or swimming while reading it.

Or having it read to you by anyone who can be finitely described. While anything describable happens before, after and during that reading.

It includes:

- all those stories, with every possible typo in every place
- ... in any language that can be expressed with a 26-character set
- ... in the style of any author, living or dead
- ... backwards and forwards, and with all of its words randomized, anagrammed and palindromed, or replaced with the word before them in the dictionary, or the word after

It includes every 2-hour episode of every mini-series, and the transcript of every reality show episode. As hard as it may be to imagine, if there's such a thing as a script so perfect that it won't receive a single note from any executive who reads it, that's in there too.

But that's it. After we've told those **S** screen stories, there are no more to tell.

Now, picture our **S** stories in alphabetical order. The first "story" will be nothing but solid repetitions of (let's say) the letter *a*, corner to corner, page to page. The last will be nothing but (let's say) semi-colons. Between those two extremes you'll find the screenplays for *Chinatown* and *The Fast And The Furious II*, along with every script you or I have ever written.

But how many of these can't be stories in any meaningful sense? How many can we summarily eliminate from **S** to chop that big number down?

### **Shaving It Down**

Mathematician G.H. Hardy (1877-1947) once estimated the number of possible chess games at 10 to the power of 10, raised to the power of 50, a far larger number than **S**. In 1950, Information Theory pioneer Claude Shannon reduced that upper bound to  $10^{43}$  by eliminating moves that couldn't arise in a legal chess game. Likewise, can't we reduce **S** by recognizing that some of our stories, by any reasonable standard, don't qualify?

The difficulty is with the word reasonable. Most writers would agree that permutations of nothing but the 9 punctuation characters may be removed

from our list of stories, in one stroke reducing **S** by  $9^{345,600}$ . (A number that, written out, has 329,000 digits.)

We might eliminate all stories which consist entirely of repetitions of a single character {*dddddd*}, including the “blank page” script, with nothing but spaces. But this cuts only 35 stories.

How about cutting any repeating cycles of characters, with *any* period {*abcdabcdabcd...*} {*joptrjoptr...*}?

Grammatical rules could be invoked to cut a wider swath. There’s a specific number, for instance, of scripts on our list that don’t end in a period.

But how small can **S** really get? After what we might call impossibilities have been eliminated, we’re still left with an incomprehensibly large number of scripts consisting of random words and pseudo-words. Information Theory speaks of the compressibility of data. The 120-page script comprising, for example, 345,600 repetitions of the letter *a* is maximally compressible. It can be fully encoded thusly: “**type the letter *a* 345,600 times.**” From those instructions alone, that “story” can be recreated.

But when we use language creatively it’s essentially random; no program can be written that significantly compresses it. From the character string {*I have of late, wherefore I know not, lost all my mirth...*}, there exists no algorithm to squeeze down everything that follows. A random string of characters is one that can’t be described more briefly than by retyping the string itself.

So: how “non-random” is a shootable screenplay, a meaningful story? If you put Ted Tally’s script for *Silence Of The Lambs* in the “non-random” pile... there are 99<sup>th</sup> percentile nonrandom “scripts” sitting alongside it that look like this: {*rtybnlopp the orpthuis....*}

We can’t even eliminate scripts containing words not found in the dictionary – every script with an “oooof!” or a “whatthefuck?” would disappear.

How about syntactic and semantic rules? Can we eliminate from **S** any script that contains the string {*the was it because*}?

Not necessarily. What if someone in your script is just reading aloud a list of words – clues, perhaps, to a murder, or to a crossword puzzle? It's difficult to justify eliminating *any* ungrammatical construction, even though we recognize that a script consisting of nothing but them isn't much of a story. (Also: that 18-character string above can appear in 345,585 different places in a 345,600-character screenplay, slightly fewer if we introduce a rule for hyphenation)

Then how do we justify leaving {*the was it because*} but eliminating {*abcdefghijklm...*}?

If we can't justify eliminating grammatically incorrect, even absurd stories from **S**... then, though common sense suggests only a vanishingly small percentage of them can ever be what we'd call "stories," I don't see how we get, and defend, a smaller number as our working definition.

I'd be interested if a reader with a mathematical bent can justify setting a smaller upper bound.

But, at any rate, let's not hear any more nonsense about there being an infinite number of stories. **S** is large enough to keep us all busy for now.

Andrew Nicholls