Through a Lens, brightly:

How the worldview of incredible fiction describes the spirit of the age

Applying digital tools to analyse a fictional genre
Science Fiction

unbounded
inclusive
immediate

“the future now”

(Ballard, 1986)
Why science fiction?

- Popular fiction reflects human interests – science fiction has become ‘normal’ fiction in demographic appeal
- A lack of content limitations in science fiction encourages exploratory thematic and mythological content as well as encouraging writing that mirrors “the future now”
- Other forms of fiction are more constrained by genre characteristics
- Magazine science fiction is rapidly produced, and ephemeral – a form of literature closely linked to contemporary society
- It includes editorial, reader feedback and advertising content that provides additional contextual information

Ben Menadue James Cook University
A standard demographic (compared to US census data)

Science fiction consumer age profiles

Current online survey built with Google Forms (n=865)

- Under 15: 0.31%
- 15 - 20: 4.34%
- 20 - 30: 23.57%
- 40 - 50: 31.32%
- 50 - 60: 26.51%
- 60 - 70: 12.25%
- 70 - 80: 1.40%
- 80+: 0.31%

Ben Menadue James Cook University
There’s a lot of it to work with...
Theoretical approach

- Synthesising quantitative and qualitative methods offers the potential to find results that demonstrate improved statistical robustness over purely qualitative approaches.

- Computer technology now exists both for quantitative analysis and the coding of qualitative data, which makes a large-scale approach feasible (and Franco Moretti goes even bigger, and broader).

- It was impossible until the latter part of the 1980s to carry out meaningful computational text analysis at all.

- Today manually attempting the scale of my work is simply unfeasible (reading one text a day would take 12 years).

Ben Menadue James Cook University
Digital data collection

- Primary literature analysis
  - Quick
  - Methodical
  - Quantitative
  - Inexpensive (if you already have the material…)

- Online surveys
  - Cheap and surprisingly easy
  - Qualitative and quantitative (Likert scale and free text answers)
Methodology “zoomed out”: the scale of the digital approach

Amazing Stories: one publication, by decades, by years, by issues, by individual pages:

- 80 years
- 960 issues
- 153,600 pages
- 46,000,000 words
Automation (Mac)
Adobe pdfs created from folders of page images
Extracting text from PDF pages using Adobe Acrobat Optical Character Recognition (OCR)

This takes quite some time, even with a fast processor and solid-state drive – approximately 80 issues a day.

In the early 80s, John Burrows took several years to process and analyse six novels by Jane Austen (Burrows, 1987).
Stripping out OCR artefacts

```python
# Processes an OCR output file
def process_ocr_output(inputFilename, outputFilename):

    # Read the raw OCR output from the input file
    raw = file_get_contents(inputFilename)

    # Strip any characters that are not alpha-numeric or whitespace
    processed = re.sub('[^a-zA-Z0-9\s]', '', raw, flags=re.IGNORECASE)

    # Convert all whitespace characters (such as newlines and tab spaces)
    # (This ensures that we preserve spaces between newlines and words)
    processed = re.sub('[
\s]', ' ', processed)

    # Remove any non-numeric single-character words, except for 'I' or 'a'
    processed = re.sub(' [^0-9ai] ', ' ', processed, flags=re.IGNORECASE)

    # Strip excess whitespace
    while (processed.find(' ') != -1):
        processed = processed.replace(' ', '')

    # Write the processed data to the output file
    file_put_contents(outputFilename, processed)
```

Courtesy of: Adam Rehn JCU School of Business (Information Technology)
Ben Menadue James Cook University
3rd November 2016
Traveler agents sore ready or traveling will sore through various duties constantilry.

Proserit in en mouths are millions of disease guns the most at first sign a trouble.

Gggglewitra full listening and keep it up by tematically also consult your doctor lis.

Uterine cold and sort throat because trade tomnion arc those of throat all it destroys.

The gtrt11s that cause them infllllinzl a a though safe to use full in for un reason.

Ball nsitalce is strength it is fatal to even nature 110 longer is able to with the staphylococcus aureus pus and their attacks go upper 2000000000 of rangmpto second.

Bacillus typhosus typhoid in counts handcaus11g disease face coiliitions that weaken.

Had resist session and government poorair sudden keepabottleoflisterinehandinhome.

Changes peratureolrexposure and and office tuck one in your bag when could not make this statement unless contact it people traveling alert coin any kills a ooo0000 from crowded cars trains and houses st louis fo a the fntistplci st a sect do ti on man covered hidden possible never vitriol hise his ll income 80801 an luuminatiog.

Message of vital interest to very au who ever yes to earn my a head see for yours of roar able volume that he men the fivefiguresalary coat by a a enable any thou years on the roadwithout collar from a to fo piece every man we were a a copy for to sales of bellltte it now free them to work for you no what your opinion is now the key to later saletanshit wm iii to a new fnsht into this fasclnatlog highly paid profession.

You tile coupon now a national salesman raining associate tin kept a792 a old chicago iii us und pnnfples salesmanslj how all can put a aifiohofrug teach j1ottotll of til filter cystic giveyou outfits of radio parts to go ev6fi had low like graduate edward.

Stand chief operator of war or frank jones 922 guadaluw st san angelo ten builder and operator of station igni and manager or the best equipped radio shop in tile southwest or help you a badly business of your own like richard butler he no with st pua pa who made around 500 a month covered with a job as motorway allen he enrolled.

The famous course that pays for tell spare time earnings are easy in radio almost from the time rou knoll page 1807 must ave nashville tenn made in his spare time.

RegEx assistance from Bruce Belson and autocorrect script by Peter Norveig
Summary
(Some) Digital Methods for Text Analysis

Quantitative analysis of content:

- Bulk extraction of text from approximately 50% of all pulp science fiction published in the 20th century – 4,000 volumes, or nearly 200,000,000 words…
- Compiling PDFs, OCR: scripting, automation and RegEx
- Text analysis – word frequencies, comparing content across time periods to highlight points of interest
- Using NVivo (which creaks) and R libraries
- Benchmarking of results against broader corpus of English text using Google Ngrams

Ben Menadue James Cook University
Limitations and caveats

- Sources restricted to US and UK periodicals from the mid-1930s until the end of the century – earlier literature tends to be based on crime and western pulp formulas.

- It’s not just about text analysis – editors, agents, writers and publishers also played a part.

- Online survey data is subject to known factors such as technology use, gender bias of survey responses and survey distribution patterns.

- Comprehending data on this scale can be challenging.
Visualising frequency results
Wordclouds: Stand on Zanzibar (1968)
Benchmarking word frequencies 1945 to 1994: Google ‘lots of books’ vs *Astounding Science Fiction*

<table>
<thead>
<tr>
<th>'45 to '54</th>
<th>'55 to '64</th>
<th>'65 to '74</th>
<th>'75 to '84</th>
<th>'85 to '94</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 one</td>
<td>one</td>
<td>one</td>
<td>one</td>
<td>one</td>
</tr>
<tr>
<td>2 time</td>
<td>time</td>
<td>new</td>
<td>new</td>
<td>new</td>
</tr>
<tr>
<td>3 new</td>
<td>new</td>
<td>time</td>
<td>time</td>
<td>time</td>
</tr>
<tr>
<td>4 man</td>
<td>man</td>
<td>life</td>
<td>way</td>
<td></td>
</tr>
<tr>
<td>5 life</td>
<td>life</td>
<td>life</td>
<td>way</td>
<td>life</td>
</tr>
<tr>
<td>6 way</td>
<td>way</td>
<td>way</td>
<td>man</td>
<td>world</td>
</tr>
<tr>
<td>7 war</td>
<td>world</td>
<td>world</td>
<td>world</td>
<td>man</td>
</tr>
<tr>
<td>8 world</td>
<td>war</td>
<td>war</td>
<td>war</td>
<td>war</td>
</tr>
<tr>
<td>9 last</td>
<td>last</td>
<td>last</td>
<td>last</td>
<td>last</td>
</tr>
<tr>
<td>10 future</td>
<td>future</td>
<td>future</td>
<td>future</td>
<td>future</td>
</tr>
<tr>
<td>11 earth</td>
<td>space</td>
<td>space</td>
<td>space</td>
<td>space</td>
</tr>
<tr>
<td>12 space</td>
<td>earth</td>
<td>earth</td>
<td>earth</td>
<td>earth</td>
</tr>
<tr>
<td>13 star</td>
<td>moon</td>
<td>star</td>
<td>star</td>
<td>star</td>
</tr>
<tr>
<td>14 stars</td>
<td>star</td>
<td>moon</td>
<td>moon</td>
<td>moon</td>
</tr>
<tr>
<td>15 moon</td>
<td>stars</td>
<td>stars</td>
<td>stars</td>
<td>stars</td>
</tr>
<tr>
<td>16 mars</td>
<td>planet</td>
<td>planet</td>
<td>planet</td>
<td>planet</td>
</tr>
<tr>
<td>17 planet</td>
<td>mars</td>
<td>mars</td>
<td>mars</td>
<td>mars</td>
</tr>
<tr>
<td>18 Relative Frequency Rank (Ngrams) 1945 to 1994</td>
<td>Relative Frequency Rank <em>Astounding SF</em> 1945 to 1994</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Frequency graphs
Google Ngrams: Atomic vs Nuclear

(Google Ngrams, 2015)
Historical context of word frequency
Ngrams: mapping to historical events

(Google Ngrams, 2015)
Outward reading: “zooming in”

“Outward reading”* of texts including themes and ideas identified in quantitative analysis which are not otherwise readily accessible:

- Editorials,
- Reader feedback surveys (quantitative and qualitative),
- Letters to the editor and advertising

A literary-critical approach

*(Scholes & Wolfman, 2010)
The long article on dianetics by L. Ron Hubbard, in this issue, is, I feel, a highly important publication indeed. The article describes a technique of mental therapy of such power that it will, I know, seem fantastic. If so, it can also be said that the power of the human mind is, indeed, fantastic. I want to assure every reader, most positively and unequivocally, that this article is not a hoax, joke, or anything but a direct, clear statement of a totally new scientific thesis.

Dr. Joseph A. Winter, M.D., was asked to write the introductory note; he has studied the techniques in detail, and learned the techniques himself. I have investigated the material myself. Dr. Winter can speak as a medical expert; I can only say that my investigations have led me to the conviction that the phenomena I have observed definitely merit publication of this material for wider analysis and testing.

It would have been wholly unfair to publish such revolutionary material until such time as Hubbard’s textbook on the technique is available. That book is in manuscript now. Modern Science of Mental Health is available now from Hermitage House. It contains the exact description of the procedures used.

Hubbard, as an engineer, has tackled the problem of the mind from the scientific method. Basically, that method is:

1. Gather all available data that is, or appears to be, relevant.
2. From the data, form an hypothesis.
3. On the basis of the hypothesis, make a prediction.
4. Experiment to check the validity of the prediction.
5. Vary the experiments, and collect more data.
6. When the new theory breaks down, take the now-collected data and formulate a new hypothesis.
7. Go back to step 3.

Most readers of this magazine are fully accustomed to that method; to those who are not, the above described process is not circular.
There are over 100 Dianetics / Scientology adverts from 1950 to the present day - frequency and content is variable

Worlds of Tomorrow v05n01 p164 Dianetics Advertising
Fantasy and Science Fiction v01n01 (December 1950)

“Vianetics: a hook review by C. DALY KING, PH.D.

THIS volume is full of assertions and claims, and frequent reference is made in it to "scientific evidence," but your reviewer could find no item of such evidence in its 400-odd pages.”

3rd November 2016
Outputs

- Synthesis of the Qualitative with the Quantitative to identify correlations and test validity
- Investigation of the world view: enduring human themes and mythologies
- Reading the spirit of the age: how people responded to content that is a specific product of their time – new and revolutionary ideas, influence of historical events and changes in popular culture and society
Questions

Yes, I’d like to hear it, Hal. Sing it for me.
References

References cited in the text
Brunner (1968), ‘Stand on Zanzibar’, Del Rey, NY.

Images
Images used are published under creative commons, fair use or copyright expired conditions unless otherwise noted below:

Space photography: NASA; public domain.
Astounding Science Fiction Covers, © Street & Smith Publications.
Book cover images all © Chris Foss www.chrisfossart.com/ except cover to 2001 A Space Odyssey and associated animated gif images, © Metro Goldwyn Meyer (1968)
Wordle (2015) http://www.wordle.net/create

Ben Menadue James Cook University