Mapping the Evolution of Legislation

a bioinformatics approach

Ruth M. Dixon and Jonathan A. Jones

University of Oxford

PSA Political Methodology Conference UCL June 2016
About thirty major pieces of government legislation are produced every year in the UK, and most are subject to hundreds, even thousands, of amendments during the parliamentary process.
Why might we wish to map this process?

Amendments are central to the parliamentary process, and can throw light on the political manoeuvring involved in the production of legislation.

For instance, Christopher Foster in ‘British Government in Crisis’ (2005) argued that legislation is increasingly poorly prepared, leading to more late-stage amendments and less parliamentary scrutiny.

Can we test this assertion?
Counting amendments...

... is possible but is very laborious and time-consuming.

There are very few quantitative studies of amendments – but see e.g. work by Amie Kreppell, George Tsebelis, Meg Russell, Lanny Martin and Georg Vanberg.
Is there another way?

**Bioinformatics** is the study of DNA sequences. DNA encodes genetic information in a four-letter ‘alphabet’ (the four bases A, C, G and T).

Bioinformatics can be used to track evolutionary relationships.

Example: mutations occurred in a gene in humans and other primates that mean that we (unlike most mammals) can’t make Vitamin C.
Bioinformatics software can handle large amounts of data.

Dark colouration of the peppered moth is caused by the insertion of 22,000 bases into a gene involved in wing development.


Photos by Olaf Leillinger (License: CC-BY-SA-2.5)
Mutation of genes and bills

Like genes, bills evolve by accumulating ‘mutations,’ that is, addition, deletion, and substitution of information.

Our method maps changes to the text of bill versions in a similar way.

Amendment of the Police Reform and Social Responsibility Bill (HoC committee)
Bill versions have a formal structure...

**A BILL**

Make provision about the administration and governance of police forces; about the licensing of, and for the imposition of a late night levy in relation to, the sale and supply of alcohol; for the repeal of sections 132 to 138 of the Serious Organised Crime and Police Act 2005 and for the prohibition of certain activities in Parliament Square; to enable provision in local authority byelaws to include powers of seizure and forfeiture; about the control of dangerous or otherwise harmful drugs; to restrict the issue of arrest warrants for certain extra-territorial offences; and for connected purposes.

BE IT ENACTED by the Queen’s most Excellent Majesty, by and with the advice and consent of the Lords Spiritual and Temporal, and Commons, in this present Parliament assembled, and by the authority of the same, as follows:—

**PART 1**

**POLICE REFORM**

**CHAPTER 1**

**POLICE AREAS OUTSIDE LONDON**

Police and crime commissioners and chief constables
...suitable for line-by-line comparison

But typeset legislation presents complexities due to

- page headers
- line and page numbers
- renumbering of sections
- front- and end-matter
- idiosyncrasies of legislative typesetting

So, the text file must be simplified before comparison.
Text simplification

The whole text is copied from the pdf into a text-editor such as Notepad, preserving line-breaks.

A Python script is used to identify and strip out:

1. line and page numbers
2. page headers
3. all remaining numbers and (most) punctuation.

Finally, front- and end- sections are removed by hand.
Text comparison

‘Simplified’ text versions are compared with (free) text-comparison software – e.g. Winmerge – and a ‘patch’ or difference file is created.
Attribution of differences to parliamentary amendments

The patch file contains some ‘spurious’ differences that were not due to amendments (and were not removed during text simplification), e.g. formatting changes and typo corrections.

These spurious differences require human intervention to identify and remove – some are difficult to classify.
Graphic display and report

Another Python script analyses the cleaned-up patch file to create the graphic display and to produce a report of additions, substitutions, and deletions.

Part of patch file

```
187,188c188,189
< A police and crime commissioner may not issue or vary a police and crime plan unless the relevant chief constable agrees to the plan or the variation
---
> A police and crime commissioner must consult the relevant chief constable before issuing or varying a police and crime plan
```

Part of Python script

```
49 with open(patchfile) as infile:
    for line in infile:
        if line[0].isdigit():
            outfile.write(line)
            line=line.rstrip()
        if 'a' in line:
            action='a'
            col='b'
            na+=1
        if 'd' in line:
            action='d'
            col='r'
            nd+=1
        if 'c' in line:
            action='c'
```
Report Output

...  
5716, 5718d6110  
5722, 5727c6114, 6115  
5740, 5742d6127  
6881a7267, 7269  
7070a7459, 7460  
8851, 8857d9240  
9028c9411, 9418  
9048a9439, 9440  
9052c9444, 9476  
9199c9623  
12 additions  
5 deletions  
57 changes  
74 total
Changes made in the House of Commons Report Stage of the Police Reform and Social Responsibility Bill (2011)
Validation

1. Automated text simplification
2. Identification of differences attributable/ not attributable to parliamentary amendments
3. Relationship of the number of text differences to the number of parliamentary amendments
1. Effect of text simplification

Text simplification progressively removes irrelevant differences

- Initial comparison of raw text from pdfs
- Line and page numbers removed
- Headers removed
- Remaining numbers and most punctuation removed
- Front and end matter removed

Commons Committee Stage

Differences detected
Effect of text simplification

Text simplification progressively removes irrelevant differences

- Initial comparison of raw text from pdfs
- Line and page numbers removed
- Headers removed
- Remaining numbers and most punctuation removed
- Front and end matter removed

Parliamentary Stages of PRSRA 2011
2. Attribute remaining differences to parliamentary amendments

Differences after automated text simplification

Differences attributed to parliamentary amendments

‘Irrelevant’ differences result from typo corrections and format changes plus a few more substantial changes
Confirm whether each difference was caused by parliamentary amendment

Parliamentary Stages of PRSRA 2011

- Commons Committee Stage
- Commons Report Stage
- Lords Committee Stage
- Lords Report Stage
- Lords Third Reading (and ping-pong)

Differences

- All differences after automated text simplification
- Differences attributed to parliamentary amendments
- Differences confirmed as due to parliamentary amendments

Attribution accuracy 97%
3. How do these difference counts relate to the **number** of parliamentary amendments?
More differences than amendments if

...a substantial block of text replaces another similar one

Replacing Schedule 15 required just two parliamentary amendments, but resulted in almost a hundred text differences:
Fewer differences than amendments if ...

...several parliamentary amendments affect the same short block of text.

Here, one deletion resulted from four parliamentary amendments:
Text changes during the parliamentary evolution of PRSRA 2011
Conclusions

• This semi-automated method accurately counts and maps changes to the text of bill versions resulting from parliamentary amendments (but does not give the exact number of amendments).
• Far quicker than counting amendments by hand.
• The patch and report files contain qualitative and quantitative information, allowing further analysis of the content, amount, and location of the amended text.
Future developments

• Extend method to older bills (need to address incomplete availability of pre-2008 versions and lower quality pdfs).
• Extend method to recent xml versions – this should allow us to remove more formatting changes automatically.

Questions to address …

• How amendment patterns vary
  • ... over time?
  • ... by government (party, size of majority, coalition/one-party)?
  • ... by policy area?
  • ... by legislature?
Biston betularia by Olaf Leillinger (Wikimedia Commons License: CC-BY-SA-2.5)