Mapping the Mutations in Legislation

Ruth M. Dixon and Jonathan A. Jones
Mapping the Mutations in Legislation

Ruth M. Dixon and Jonathan A. Jones

University of Oxford

Author for correspondence: Ruth.dixon@bsg.ox.ac.uk

Paper prepared for Political Studies Annual Meeting Glasgow 2017

This study is a work-in-progress, and comments and suggestions are welcomed. Please check with the authors before citing or quoting this paper.

Abstract

One of the central roles of legislatures is to scrutinize and amend legislation, but little academic attention has been given to the level of amendment which draft laws receive. It is not known, for instance, how much of the text is amended during the passage of a typical bill. This question is of more than academic interest. While draft legislation is subject to pre-legislative consultation, amendments introduced during the parliamentary process may receive rather less scrutiny. This paper describes a novel method to map and record the changes that take place in UK legislation as it passes through the parliamentary process. In a pilot study of seventeen UK government bills from 2009 to 2013, we found that on average almost thirty per cent of the lines of legislative text were amended (running in many cases to thousands of altered lines of text) and bills increased in length by an average of over fifteen per cent. Whether this degree of alteration is ‘too much’ is beyond the scope of this study, but it demonstrates the extent to which the final legislation may differ from the initial draft.

Acknowledgements

We are very grateful to the Legislation Offices of the House of Lords and House of Commons for providing us with bill texts in unpublished formats. We also thank several former and current parliamentary officials and civil servants for generously and patiently answering our questions about the parliamentary process.
1. Introduction

An important role of Parliament is to scrutinize and amend legislation. In the UK, about forty major pieces of primary legislation are produced each year. Typically, hundreds of government and opposition amendments to each bill are proposed at each amending stage in the two legislative chambers (the House of Commons and House of Lords), and almost all proposed government amendments are ‘agreed’ (that is, incorporated into the subsequent version of the bill). Such amendments serve a variety of purposes. Government amendments may reflect changes of policy and circumstances that arose since the bill was drafted. They may arise from earlier rounds of consultation and debate, or be versions of amendments originally proposed by opposition or backbench legislators (Russell 2013:168). And some may be needed to repair defective legislation, too hastily put together (Foster 2005). Others are known as ‘technical’ or ‘drafting’ amendments, for instance making changes consequent on other amendments.

We do not claim that counting amendments or measuring the degree of change of legislative text tells us a great deal about the parliamentary process. Detailed studies of specific amendments allow scholars to track how policies change during the process of legislation, or to understand the political purpose of particular changes (e.g. Kreppel 1999; Russell 2013; Tsebelis et al. 2001). Case studies of particular bills (e.g. Bovey 2008; Timmins 2012) or parliamentary roles (e.g. Page 2003; Page 2009; Levy 2010) illuminate the process in a way that a quantitative analysis cannot. But we suggest that a quantitative measure of the degree of amendment also highlights an important aspect of the parliamentary process, and complements other types of analysis.

But how much of the bill text is altered through amendment? A study of UK bills on two legislative topics (Health and Criminal Justice) since 1972 showed that on average over four hundred government amendments were agreed for each bill (Hood and Dixon 2015: 211). Existing quantitative studies of legislative amendments are few in number and generally include only a limited number of pieces of legislation. The study mentioned above included twenty-nine UK bills.

1 UK legislation is referred to as a bill as it proceeds through parliament and an Act of Parliament when it receives Royal Assent. In the UK legislative process, amendments can be made in five Parliamentary stages: the House of Commons Committee and Report stages, and the House of Lords Committee, Report and Third Reading stages.
Amy Kreppel (1999) studied twenty-four EU legislative proposals. An exception is George Tsebelis and colleagues (2001) who hand-coded some 5000 amendments in 231 EU legislative proposals. A few studies provide an overview of legislative activity, such as Gavin Drewry and Jenny Brock’s analysis of the amendments made in the UK House of Lords in one parliamentary session (Drewry and Brock 1993). This near-absence of quantitative research is partly because the process is so laborious. In the UK, agreed amendments are recorded in a different way at each parliamentary stage. There is also little administrative data available: for instance the House of Lords Public Bill Office provides statistics on the number of amendments made in that chamber, but there is no equivalent information from the House of Commons. Furthermore, a count of amendments does not measure the proportion of text changed.

The advent of advanced data-processing methods suggests that there should be a way of automating, or at least streamlining, this type of analysis. In other fields, such as bioinformatics or software version control, techniques for comparing ‘code’ (whether genetic code or lines of software) are well established. Since DNA sequencing was developed in the 1970s (Sanger et al. 1977) the field of bioinformatics has used such comparisons to analyse evolutionary relationships and the changing functions of genes (e.g. Emms and Kelly 2015; van’t Hof et al. 2016). Legislative changes can be likened to the mutations in DNA, which involves additions or deletions of sections of code, and single word (or letter) changes. Our methodological choices, however, were determined by the structure of legislative text. Genetic code is compared on a letter-by-letter basis, but legislative texts are better compared line-by-line or paragraph-by-paragraph. This is more akin to software version control (see e.g. Kemper and Oxley 2012), and so our method uses software designed for line-by-line comparisons.

Taking inspiration from the advances in other fields, we accordingly developed a method for mapping the amendments agreed in successive parliamentary stages during the passage of a piece of legislation. This paper describes our method and its validation, as well as our preliminary results. Finally, we discuss the implications of our results and describe future directions for this research.
2. Methodology and Data Sources

2.1 A Semi-automated Approach to Mapping Amendments

Our method is outlined in Figure 1.

Figure 1. Method outline. (Step 3 is only required during method development.)

1. Obtain text of bill versions, strip out extraneous material
2. Line-by-line comparison to generate a ‘patch file’ of differences
3. Validate differences (amendments?)
4. Analyse differences to create data-visualisation and quantitative summary

2.2 Availability of bill versions

The first step is to obtain the relevant bill versions in electronic form. Up to six versions of each piece of UK legislation are published. For bills introduced to the House of Commons, these are (i) the bill as introduced, (ii) as amended in the Commons public bill committee (PBC), (iii) as amended at Commons report stage and introduced to the House of Lords, (iv) as amended in the Lords Committee stage, (v) as amended in Lords Report Stage, and (vi) the final Act of Parliament (which may have been amended in the Lords Third Reading stage and during ping-pong\(^2\)). For bills introduced to the House of Lords, the order of Commons’ and Lords’ stages is reversed, and ping-pong similarly occurs at the end of the process. If no amendments are made at a particular stage, new versions are only published when the bill enters the next chamber and when the final Act of Parliament is agreed.

Since 2007, each piece of legislation has a page on the parliament.uk website, with links to PDFs of all bill versions and to the final Act of Parliament publicly available. Before that date, most House of

\(^2\) ‘Ping-pong’ refers to a bill being shuttled between the two chambers until the exact wording is agreed.
Commons bill versions are available via http://parlipapers.proquest.com/parlipapers (for registered users), but House of Lords versions are not available online. Disappointingly, however, PDFs from before 2007, where available, are difficult to analyse by any automatic method. Early files suffer from optical character recognition errors as the PDF is generated from scanning a paper document. These documents are also structured in a way that makes separating the text from extraneous material such as marginal notes almost impossible. For this study, therefore, we worked only with bills published after 2007, which are generated from electronic sources.

As PDFs are not ideal for our method, as discussed below (section 2.5), we also investigated the possibility of using the underlying files from which the PDFs are produced. UK legislation is currently written using Adobe FrameMaker. The mark-up language (XML) extracted from FrameMaker files requires far less pre-processing than PDFs before comparison. However, these files are not publicly available.³

From 2016, XML files of some bill versions became available on the Parliament website. Final versions of UK legislation are available in XML format from legislation.gov.uk.⁴

2.3 Text simplification

In order to extract the legislative text for comparison, the complete text from PDF originals was copied into a plain text file. Material extraneous to the legislative text was removed as follows. For text derived from PDFs, a Python script (version 2.7, Python Software Foundation, python.org) was used to remove page and table headers (using a look-up file of lines to be removed)⁵, line and page numbers, and all other digits and most punctuation. Front- and end-matter were removed

---

³ We are grateful to the House of Lords and House of Commons Public Bill Offices for providing some FrameMaker files for this study.
⁴ http://www.data.parliament.uk/dataset/53 and https://data.blog.gov.uk/2010/07/29/legislation-gov-uk-api/ We have not yet included files from these sources in our study.
⁵ The look-up table was generated (partly manually) from the table of contents and other information in the bill text.
2.4 Text comparison

The resulting 'simplified' texts of pairs of bill versions were compared using the free software ‘WinMerge’ (WinMerge 2.14.0, winmerge.org) which compares text on a line-by-line basis. A patch file (file of differences) was created for each comparison which contained the line numbers of any differences, classified by WinMerge as ‘additions’, ‘deletions’ or ‘changes.’ The patch file also contained the actual text of each line that was added, or deleted, or changed.

2.5 Validation exercise: classification of detected differences

Certain types of change can occur between bill versions which do not require parliamentary amendments but are detected by our method alongside actual amendments. Such alterations are inserted by the legislation offices or arise during typesetting. These include typographical corrections, alterations to section titles, and minor changes to grammar and wording. In addition, for text extracted from PDFs, ‘formatting’ differences occur when paragraph references change after the insertion or deletion of earlier paragraphs, if words within a paragraph break differently across lines, and if tables split across more than one page. XML-derived text avoided most of these formatting problems.

In order to determine the level of changes that were (or were not) caused by amendments, we analysed over 70 bill comparisons in detail. For this exercise, we compared each successive version of 17 government bills published between 2009 and 2013. For these comparisons, each detected change was manually classified as either the result of (one or more) parliamentary amendments or as a ‘formatting’ or other irrelevant change. Where the classification was uncertain it was checked by reference to the online record of agreed amendments. After classifying each change in a

---

6 FrameMaker-derived XML files contained no line or page numbers, page headers, tables of contents, etc. Paragraph references and most other cross-references were automatically removed and the text of each paragraph appeared on a single line.

7 Some cross-references in XML files were in plain text (so were not removed during text simplification). These cross-references, if altered between versions, resulted in detected differences.
spreadsheet, an edited version of the patch file was saved which contained only the changes classified as amendments. Both the ‘raw’ and the ‘edited’ patch files were analysed as described in section 2.6.

2.6 Graphic Display and Analysis

The text comparison and validation steps resulted in two patch files for each pair of bill versions compared. The first patch file contained all changes detected by the file comparison software, and the second was an edited version containing only the changes classified as being due to amendments as described in section 2.5. A Python script was used to analyse each file to determine the number of changes and the number of lines of text amended. Comparing ‘all’ and ‘amendment-only’ changes allowed the calculation of the amount of change for other (non-amendment) reasons. The Python script also generated a visual display of the changes between the two bill versions (Figure 2). The output of the Python script was further analysed to calculate the proportion of text changed, and the number of changes detected per 100 lines of text. The percentage of lines changed between two versions was calculated as (lines deleted + lines added)/(lines deleted + lines added + lines unchanged) x 100 per cent. The number of changes was calculated from the Winmerge patch file counts as ‘additions’ + ‘deletions’+ 2 x ‘changes’ (‘changes’ always contained an addition and a deletion). The size of each addition, deletion, and change could be any whole number of lines. The file length was the length of the text file after the text simplification step (essentially the number of lines of legislative text, excluding front- and end-matter).
Figure 2. Amendments in the House of Commons Committee Stage of the Police Reform and Social Responsibility Bill 2011.

3. Results

3.1 Validation exercise: classification of changes

The results of the manual classification of changes as ‘amendments’ and ‘other’ are shown in Table 1 and Figure 3.

Table 1. Changes due to amendments and for other reasons (n=72 comparisons)

<table>
<thead>
<tr>
<th></th>
<th>Amendments</th>
<th>Other reasons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Range</td>
</tr>
<tr>
<td>Percentage of lines changed</td>
<td>7.5</td>
<td>(0 – 41.5)</td>
</tr>
<tr>
<td>Number of changes detected in the bill</td>
<td>90</td>
<td>(0 – 1555)</td>
</tr>
<tr>
<td>Number of changes per 100 lines of text</td>
<td>1.0</td>
<td>(0 – 7.4)</td>
</tr>
</tbody>
</table>
Table 1 shows that a considerable number of ‘other’ changes persist in the legislative text even after the text simplification step. However, the percentage of text altered by ‘other’ changes is relatively constant, while the percentage altered by amendments varies greatly, as shown in Figure 3. Thus omitting the ‘classification’ step would only slightly overestimate the degree of amendment for most bill comparisons.\(^8\)

**Figure 3. Percentage of text changed in 72 bill stage comparisons (ordered by overall level of change).**

![Figure 3. Percentage of text changed in 72 bill stage comparisons](image)

To test the method on XML files, we obtained all seven\(^9\) versions of the Health and Social Care bill 2012 in both PDF and XML, and conducted the same analysis on both types of file. Although the

---

\(^8\) Two comparisons contained a relatively large proportion of ‘other’ changes (over 6 per cent of the text). One was a very short bill, where the conventional change to the formatting of the long title in the final Act altered over five per cent of the text. The other was a bill that containing a large number of tables, which were picked up inconsistently when copied from the PDF originals. Both of these types of problems would be mitigated by using XML originals.

\(^9\) Unusually, this bill went through two Commons public bill committee (PBC) stages, resulting in seven versions. The tortuous passage of this bill was described in the report *Never Again?* (Timmins 2012).
PDF-derived texts contained almost twice as many lines as their XML-derived counterparts. The results from the two types of file were very similar, as shown in Figure 4.

Figure 4. Comparisons of XML and PDF originals (Health and Social Care Act 2012).

3.2 Comparisons of bills and bill stages

Turning to our pilot study of bills and stages, the results from 17 government bills (2009 to 2013) are shown in Tables 2 and 3. In Table 2, the first version of each bill was compared with the final Act of Parliament (encompassing the entire parliamentary process), and in Table 3 successive pairs of bill versions were compared for each parliamentary stage.

---

10 The difference in length arises because text copied from PDFs separates paragraphs that occupy more than one line on the typeset page into the corresponding number of lines of plain text, while XML-derived text puts each paragraph on a single line.
Table 2. Comparison of the first and final versions of 17 government bills (2009-2013).

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial length of bill text</td>
<td>7045 lines</td>
<td>(37 – 34379)</td>
</tr>
<tr>
<td>Final length of bill text</td>
<td>8202 lines</td>
<td>(39 – 34355)</td>
</tr>
<tr>
<td>Change in length</td>
<td>+15.8 %</td>
<td>(0 – 63.5)</td>
</tr>
<tr>
<td>Detected changes (all)</td>
<td>387 changes</td>
<td>(4 – 2233)</td>
</tr>
<tr>
<td>Proportion of lines changed</td>
<td>29.8 %</td>
<td>(2.0 – 74.2)</td>
</tr>
</tbody>
</table>

Table 3. Percentage of lines changed at each parliamentary stage for 17 government bills.

<table>
<thead>
<tr>
<th>Bills starting in House of Commons (n = 11)</th>
<th>Mean %</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commons Public Bill Committee</td>
<td>10.8</td>
<td>(0 – 40.7)</td>
</tr>
<tr>
<td>Commons Report Stage</td>
<td>6.5</td>
<td>(0 – 25.5)</td>
</tr>
<tr>
<td>Lords Committee</td>
<td>7.8</td>
<td>(0 – 41.5)</td>
</tr>
<tr>
<td>Lords Report Stage</td>
<td>7.2</td>
<td>(0 – 23.9)</td>
</tr>
<tr>
<td>Lords Third Reading (and ping-pong)</td>
<td>2.2</td>
<td>(0 – 6.9)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bills starting in House of Lords (n = 6)</th>
<th>Mean %</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lords Committee</td>
<td>7.0</td>
<td>(0 – 21.0)</td>
</tr>
<tr>
<td>Lords Report Stage</td>
<td>6.3</td>
<td>(0 – 24.4)</td>
</tr>
<tr>
<td>Lords Third Reading</td>
<td>3.3</td>
<td>(0 – 14.9)</td>
</tr>
<tr>
<td>Commons Public Bill Committee</td>
<td>3.5</td>
<td>(0 – 13.1)</td>
</tr>
<tr>
<td>Commons Report Stage (and ping-pong)</td>
<td>4.1</td>
<td>(0 – 8.5)</td>
</tr>
</tbody>
</table>
As Tables 2 and 3 demonstrate, this sample of government bills was subject to a considerable level of amendment. On average, almost thirty per cent of the lines in each bill were altered during the parliamentary process (this proportion was very similar whether the bills were introduced in the Commons or in the Lords). There was a wide variation between bills, from only two to over 70 per cent of lines changed.\(^1\)

Another striking finding was that the legislation grew considerably length during the parliamentary process, averaging over 15 per cent increase in length. No bill experienced an overall reduction in length during its passage through parliament. While a few bills were shortened during individual parliamentary stages, any reductions were more than made up for in other stages.

In individual parliamentary stages, up to forty per cent of the text of each bill in this sample was altered (the average proportion, across all stages, was 7.5 per cent). The level of amendment differed between stages, which is to be expected, given the different conventions and rules that apply to each stage of the parliamentary process (House of Lords 2011). For instance, issues already debated and decided cannot be reopened at the Lords Third Reading stage. Amendments made at that stage are generally intended to tidy up the bill, or for ministers to fulfil commitments made at earlier stages in the process. Accordingly, we found that the least amount of text changed at the Lords Third Reading stage. We found that the percentage of any bill version changed by amendments or other changes was independent of the overall bill length, as shown in Figure 5.

\(^1\) The very highly amended legislation was the Constitutional Reform and Governance Act 2010. This bill was unusual in that a large number of clauses were inserted during the Commons Committee stage, and then a further large number of (other) clauses were deleted by the House of Lords, shortly before the General Election in May 2010. There being no further parliamentary time available for the House of Commons to consider (and potentially reverse) the Lords’ amendments, the then Home Secretary Jack Straw conceded the amendments to allow the bill to be passed. Hansard records that Straw undertook to bring back the provisions should the government be re-elected (it was not).
Figure 5. Degree of text alteration in parliamentary stages by amendment or for other reasons, versus overall length of legislative text.

4. Discussion

While the development of our ‘semi-automated’ method for assessing the level of amendment of legislation appears (and was) labour-intensive, the time-consuming part of this study was the validation step, in which each detected change was classified as being caused either by parliamentary amendment or for other reasons such as typographical corrections. As shown above, the validation exercise demonstrated that the level of ‘other’ changes was relatively low and consistent across bill version comparisons. The vast majority of alterations could be attributed to parliamentary amendments. Future studies will therefore be able to largely omit the arduous classification step.

As we have shown in this pilot study, UK legislation changes considerably during the legislative process. The parliamentary process as a whole altered, on average, almost thirty per cent of the
lines of text, and added over fifteen per cent to the length of each bill in our sample.\textsuperscript{12} The salience, or legal effect, of these changes is beyond the scope of this study, but it is apparent that legislators are not only expected to scrutinize the first draft of each bill, but to take on board often radically changed text in successive versions. Former Parliamentary Counsel Daniel Greenberg argued that legislation does not receive the detailed scrutiny that it deserves (Greenberg 2016). As there is often insufficient time even to consider the whole text of bills, government amendments are often agreed \textit{en masse} with little or no discussion. Alterations, therefore, can bypass some of the intended parliamentary scrutiny. Amendments also avoid pre-legislative scrutiny (the period of consultation that precedes the parliamentary process) which necessarily considers only the initial version of the bill. Of course, lobbying also takes place throughout the passage of the bill, but there is often limited time to consider and respond to large numbers of proposed amendments.

High levels of amendment may also be evidence for the ‘legislative hyperactivity’ noted by Lord Bingham (2007). Although efforts are made to reduce the amount of legislation (such as the consolidating and repealing work of the Law Commission) over 30,000 ‘legislative effects’ are reported to result every year from new primary and secondary legislation, and over 4,000 new criminal offences were created between 1983 and 2009 (Office of the Parliamentary Counsel 2013). UK legislation has been described as complex and inaccessible (House of Commons Political and Constitutional Reform Committee 2013). Legal theorist Lon Fuller noted that one of the attributes of good legislation is that it should be stable enough to guide action (Fuller 1969:80). If legislation is wildly unstable even as it passes through parliament, the resulting law may not survive long before it is challenged in the courts or needs to be revised by subsequent legislation. The high levels of ‘churn’ in government, whether the rapid turnover of ministers, senior civil servants, departments, agencies, performance targets, financial data series, or policies, have been widely discussed (see e.g. Berlinski et al. 2012; Theakston and Fry 1989; White and Dunleavy 2010; Elston 2013; James 1994; Martin 2009; Soroka et al. 2009; Norris and Adam 2017). We might now add ‘bill text churn’ to this list.

\textsuperscript{12} We intend to include a larger sample of bills in an extension of this study, alongside analysis of recently available XML versions.
But do these amendments improve legislation? While that question can’t be answered by this study, our findings offer some directions for future research. Of course, the process of amendment is a necessary and valuable function of parliament, allowing legislators to correct and improve legislation after detailed scrutiny. Although the vast majority of agreed amendments are ‘government’ amendments, the scarcity of agreed ‘non-government’ amendments does not necessarily imply that governments are unresponsive to the arguments of opposition or backbench legislators. As Meg Russell showed, more than half of agreed ‘policy-type’ government amendments in twelve bills in the House of Lords were versions of amendments originally proposed by non-ministerial peers (Russell 2013). And Louise Thompson found that ministers were almost twice as likely to give an undertaking to consider opposition or backbench amendments (with a view to bringing them back as government amendments at a later stage) in committees that heard oral evidence compared to those that did not (Thompson 2014). However, the possible link between high levels of amendment and lower quality of legislation is certainly worthy of further study. A House of Lords report noted in 1987: ‘Government legislation has been more heavily amended in the Lords than in previous Sessions [...] which suggests than legislation may have been introduced without adequate consideration.’ And former government adviser Christopher Foster argued that legislation is increasingly poorly prepared and requires numerous amendments to ‘repair’ it during the parliamentary process (Foster 2005). Alternatively, it is possible that changes to parliamentary processes, such as advances in electronic drafting and dissemination of bill texts, have improved the quality of both drafting and scrutiny (Hood and Dixon, 2015: 155-156).

By providing a convenient way of measuring the degree of amendment of legislation, our methodology can therefore contribute to studies of the quality and characteristics of legislation, alongside other methods such as linguistic and mathematical analysis (e.g. Williams 2016; Bommarito and Katz 2010; Waltl and Matthes 2014) and detailed qualitative or quantitative case studies. The method also produces a file of all the lines of altered text which can be used for further analysis of the substance of the amendments. This approach can be applied to legislative texts from other jurisdictions, wherever suitable electronic versions of bill texts are available. Our study of UK

\footnote{For example, government amendments comprised 99.5 per cent of agreed Commons amendments, and 91 per cent of amendments agreed in the Lords in 29 Health and Criminal Justice Bills from 1972 to 2012 (Hood and Dixon 2015:210).}

\footnote{The House of Lords Report on the Working of the House (HL 9, 1987-88).}
legislation was limited by the lack of bill versions from before 2007, but the availability of XML versions of recent legislation suggests that studies of this type will become much more practicable and widespread in the future.

References


James, Oliver. 1994. ‘The UK Core Executive’s Use of Public Service Agreements as a Tool of Governance’, Public Administration 82 (2): 397–419.


Norris, Emma, and Robert Adam. 2017. All Change: Why Britain is so prone to policy reinvention, and what can be done about it. London: Institute for Government.


