

## The Victorian Meme Machine: Remixing the Nineteenth-Century Archive

Bob Nicholson

History has not been kind to Victorian jokes. While the great works of nineteenth-century art and literature have been preserved and celebrated by successive generations, even the period's most popular gags have largely been forgotten. In the popular imagination, the Victorians have long been regarded as terminally humourless; a strait-laced society who, in the words of their queen, were famously 'not amused'.<sup>1</sup> And yet, millions of jokes were written during the nineteenth century. They were printed in books and newspapers, performed in theatres and music halls, and retold in pubs, offices, taxicabs, schoolrooms, and kitchens throughout the land. Like many other forms of ephemeral popular culture, the majority of these jokes were never recorded and have now been lost. As a result, historians have tended to focus on weightier and more enduring forms of Victorian comedy: the literary humour of Charles Dickens, Jerome K. Jerome, Thomas Hood, Douglas Jerrold, and Oscar Wilde; the political satire of *Punch*; or the songs of the late-Victorian music hall.<sup>2</sup> This is important work, but it covers a small fraction of the period's comic output. If we want to understand the impor-

---

Unless otherwise stated, all hyperlinks in the body of the article were accessed on 5 October 2015.

<sup>1</sup> It is possible that Victoria never actually uttered this phrase — its attribution has been the subject of debate for more than a century. The *Yale Book of Quotations* (New Haven: Yale University Press, 2006) attributes it to her via an anecdote printed in an American paper (p. 789). On the other hand, Princess Alice claimed in a 1977 interview with the BBC that her grandmother had 'never said it'. However, even if it is apocryphal, the phrase has since become deeply embedded in popular understanding of the period. For more discussion of this, see Duncan Marks, 'We ARE Amused! The Comical Uses and Historical Abuses of Queen Victoria's Infamous Reproach', in *History and Humour: British and American Perspectives*, ed. by Barbara Korte and Doris Lechner (Bielefeld: transcript Verlag, 2013), pp. 133–50.

<sup>2</sup> For a sample of this scholarship, see *The Victorian Comic Spirit: New Perspectives*, ed. by Jennifer A. Wagner-Lawlor (Aldershot: Ashgate, 2000); Roger B. Henkle, *Comedy and Culture: England 1820–1900* (Princeton: Princeton University Press, 1980); Richard D. Altick, *Punch: The Lively Youth of a British Institution 1841–1851* (Columbus: Ohio State University Press, 1996); and Patrick Leary, *The Punch Brotherhood: Table Talk and Print Culture in Mid-Victorian London* (London: British Library, 2010). For key texts on Victorian music hall, see *Music Hall: The Business of Pleasure*, ed. by Peter Bailey (Milton Keynes: Open University Press, 1986); Jacqueline S. Bratton, *Music Hall: Performance and Style* (Milton Keynes: Open University Press, 1986); Dagmar Kift, *The Victorian Music Hall: Culture, Class, and Conflict* (Cambridge: Cambridge University Press, 1996); Barry J. Faulk, *Music Hall and Modernity: The Late-Victorian Discovery of Popular*

tance of humour within Victorian Britain, and to unpack the social, cultural, and political work that it performed, then we need to find a way to recover some of these long-forgotten jokes and open them up to scholarly analysis. This is a challenging proposition, but not an impossible one. Millions of puns, gags, and comic sketches *have* been preserved — often by accident — in archives of nineteenth-century print culture. Some appear in dedicated joke books and comic periodicals, but most have survived as stowaways in the margins of other texts. They are scattered throughout thousands of Victorian books, newspapers, magazines, and periodicals. While some were organized into clearly demarcated collections, others were used more haphazardly as column fillers or sprinkled randomly among other titbits of news and entertainment. Until recently, the only way to locate these scattered fragments amidst the ‘vast terra incognita’ of Victorian print culture was to identify a promising host text and then browse through it manually.<sup>3</sup> The digitization of Victorian print culture has opened up new possibilities for this kind of research. However, as this article argues, the structure of these new archives continues to bury jokes among millions of pages of other content. In order to make these, and other marginalized texts, more visible, we need to rethink the organization of our digital collections and open up their contents to creative forms of archival ‘remixing’.

In the summer of 2014, I began working with the British Library Labs on a project that aims to find Victorian jokes in their digital collections and, for better or worse, bring them back to life. The Labs is an initiative funded by the Andrew W. Mellon Foundation that aims to help researchers, developers, and artists access the British Library’s digital collections and use them in creative new ways. They run an annual competition that invites people to pitch new ideas for projects that make use of the library’s data sets. I was fortunate enough to win the competition in 2014, which allowed me to work with the Labs team for six months.<sup>4</sup> We began with two key aims:

- a) To build a high-quality, open access research database of one million Victorian jokes. These jokes would be extracted from the library’s existing digital collections.
- b) To share these jokes with modern audiences in creative new ways, including the use of images, videos, performances, and social media.

---

*Culture* (Athens: Ohio University Press, 2004); and Len Platt, *Musical Comedy on the West End Stage, 1890–1939* (Basingstoke: Palgrave Macmillan, 2004).

<sup>3</sup> Patrick Leary, ‘Victorian Studies in the Digital Age’, in *The Victorians since 1901: Histories, Representations and Revisions*, ed. by Miles Taylor and Michael Wolff (Manchester: Manchester University Press, 2004), pp. 201–14 (p. 206).

<sup>4</sup> Bob Nicholson, ‘British Library Labs: Submitted Entry for 2014 Competition’, <<http://labs.bl.uk/The+Victorian+Meme+Machine>> [accessed 4 October 2015].

The project is still in its infancy. We have made steady progress over the last year and have developed some rudimentary prototypes, but there is still a long way to go before the archive will be ready to launch. At present, our database contains 1500 jokes and has a basic search feature. As a result, this article is a reflection on work that is still in progress; it explains the rationale behind the project and outlines the work that we have undertaken so far. Part one explains why Victorian jokes are worthy of academic attention and demonstrates how the most laboured of puns can reveal new insights into nineteenth-century culture and society. Part two explores the relationship between Victorian jokes and existing digital archives, and considers the pros and cons of liberating them from the restrictions imposed by these collections. Finally, part three documents the progress we have made so far. In particular, it reflects on the development of our ‘Victorian Meme Machine’ tool. This piece of software takes textual jokes and automatically pairs them with a random image drawn from the British Library’s digital collections. These ‘remixed’, visual versions of the jokes are designed to be shared using modern social media platforms such as Twitter, Facebook, and Tumblr. The process is kick-started by our ‘Mechanical Comedian’ tool, which uploads a random joke to Twitter every lunchtime, and it is up to the public to decide whether a Victorian joke goes viral or slips once again into obscurity.

### The value of Victorian jokes

Before we discuss *how* Victorian jokes might be rescued from the enormous condescension of posterity, it is important to consider *why* they are worth saving. In order to do this, we need to unpack their relationship with three different audiences: the Victorians who originally wrote, told, and consumed the jokes; the historians who use them to explore Victorian culture and society; and, finally, the modern-day artists and audiences who might perform and consume them anew.

Firstly, it is important to address the ‘not amused’ myth head on. The idea that the Victorian era was humourless, though long since challenged by historians, remains in popular circulation and can sometimes underpin the assumptions of even the best academic work in the field. As recently as 2006, for example, Vic Gatrell concluded his groundbreaking study of sex and satire in Georgian London with the ‘watershed’ decade of the 1820s. After this, he argues, the bawdy, bodily humour and racy satire of the Georgian period was silenced for a hundred and fifty years by ‘respectable’, ‘squeamish’, ‘fastidious’, and ‘moralizing’ members of the ascendant Victorian bourgeoisie.<sup>5</sup> Changes in public manners and morality did

<sup>5</sup> Vic Gatrell, *City of Laughter: Sex and Satire in Eighteenth-Century London* (London: Atlantic, 2006), p. 19.

indeed take place during the early decades of the nineteenth century, but humour continued to occupy an important position within Victorian culture and society. The uninhibited ‘bum and fart’ jokes of Gatrell’s caricatures slipped (though never entirely) out of vogue, but they were replaced by other forms of popular humour that circulated widely and performed equally important cultural work.<sup>6</sup>

A comprehensive survey of these developments lies beyond the scope of this article, but a few trends and landmarks are worth noting. The pun — arguably the Victorians’ best-loved genre of joke — was theorized as early as 1826 with the publication of C. M. Westmacott’s *Punster’s Pocket-Book*.<sup>7</sup> This volume contains comic essays and poems on the ‘art of punning’, and includes an anthology of choice puns supposedly made by well-known aristocrats, politicians, writers, and other public figures. Throughout these anecdotes, the demonstration of wit and a willingness to crack a joke is celebrated as a positive character trait. For example, in the introduction to one particular section, Westmacott salutes ‘the pleasant, punning, conversational powers’ of a friend whose ‘whim, wit, and great good nature [were] not more esteemed, than his unaffected manners, and sincerity of disposition justly entitle[d] him to’ (p. 125). The pun, in other words, was a signifier of sociability and a key component in the ‘table talk’ of Victorian society. This was not a one-off publication, but an early example of a comic genre that continued to flourish throughout the nineteenth century. Anthologies of puns, jokes, and witty remarks made by famous figures continued to be published during the period. Other notable examples include the *Railway Book of Fun* (c. 1875), which instructed its readers that ‘cheerfulness [was] a Christian duty’ and a vital component of good mental and physical health.<sup>8</sup> Crucially, the gags in these collections were not simply designed to be consumed in silence during a long railway journey; they were read aloud to fellow passengers, retold to friends in the pub, and creatively repurposed to suit new audiences. For example, a reviewer of a joke book published in the 1860s concurred with its author that “the most fitting place for [this] book [was] in the hands of the young gentleman who has undertaken to amuse an assembled party” brought together [...]

<sup>6</sup> In her study of Victorian pornography, Lisa Sigel complicates Gatrell’s conclusions by demonstrating that the body remained a source of humour in the nineteenth century. See Lisa Z. Sigel, *Governing Pleasures: Pornography and Social Change in England, 1815–1914* (London: Rutgers University Press, 2002).

<sup>7</sup> Bernard Blackmantle, *The Punster’s Pocket-Book; or, The Art of Punning* (London: Sherwood, Gilbert, and Piper, 1826) <<https://archive.org/details/punsterspocketbooowestiala>> [accessed 5 October 2015]. Westmacott published the book under the pseudonym of Blackmantle, though modern reprints often use his real name.

<sup>8</sup> Richard Brisk, *The Railway Book of Fun* (London: Nicholson, [n.d.]), p. 5 <<https://archive.org/details/railwaybookoffunoobris>> [accessed 5 October 2015].

for the purpose of spending a pleasant evening'.<sup>9</sup> Sadly, we know little about the oral second life of these jests, or indeed the jokes that were performed each night by professional comedians in the country's circus tents, theatres, and music halls. Jacky Bratton and Ann Featherstone's study of the personal gag book of a Victorian clown provides a rare glimpse into this ephemeral world and demonstrates the central role that joke-telling played in the orchestration of circus entertainment.<sup>10</sup>

These books and public performances are simply the tip of the iceberg. The popularity and significance of jokes during the Victorian period is best demonstrated by analysing their presence in the popular press. Jokes began to appear sporadically in British newspapers as early as the 1830s. A particularly interesting example comes from the *Northern Liberator*, a short-lived mouthpiece of the Newcastle Chartists, which briefly published joke columns that had been imported from America.<sup>11</sup> However, during the second half of the nineteenth century, weekly columns of British and American jokes became a core component of the so-called 'new journalism'. They appeared across a wide range of publications, from mass-market papers like *Tit-Bits*, to women's domestic magazines, boys' story papers, and the back pages of hundreds of provincial weeklies. These columns usually included between twenty and thirty jokes in each instalment and ran for years at a time without interruption; over the course of a decade, a single paper was capable of publishing more than ten thousand jests. It is important to stress that editors usually clipped many of the jokes from publications such as *Punch*, and that we should therefore not expect to find tens of thousands of original gags in each provincial paper. Nevertheless, our initial survey of the British Library's digital newspaper holdings suggests that hundreds of thousands of Victorian jokes are waiting to be uncovered. This makes them the most promising source of material for our Victorian jokes database, although extracting them from newspapers poses a series of technical challenges that will be analysed in subsequent sections of this article. For the purposes of our current discussion, the presence of jokes in so many nineteenth-century periodicals acts as a final rebuttal to the 'not amused' myth. They became a staple feature of the popular press at a time when its editors were operating under growing commercial pressures and were always searching for innovative new ways to attract and retain a mass readership. *Tit-Bits*, one of the most successful and influential players in

<sup>9</sup> 'Illustrated Gift-Books For The Young', *Morning Post*, 5 December 1867, p. 3. The book under review was Hugh Rowley's *Puniana*, which is discussed in more detail in a subsequent section of this article.

<sup>10</sup> Jacky Bratton and Ann Featherstone, *The Victorian Clown* (Cambridge: Cambridge University Press, 2006).

<sup>11</sup> A column of American jokes appeared in the paper each week between 7 September 1839 and 26 October 1839 under the titles 'Whims of Jonathan' and 'Scraps from the Far West'.

this emerging market, consistently filled its front cover with jokes and scattered them throughout its other pages as column fillers. Similarly, *Lloyd's Weekly News* introduced a regular column of imported American jokes in 1896 — the same year that it became the first newspaper in the country to achieve a circulation of one million. Countless provincial papers adopted similar strategies as they shifted their focus from news to entertainment. For example, in 1883 the *Hampshire Telegraph* moved away from its traditional focus on naval news and shipping intelligence, and sought to broaden its readership through the introduction of a magazine-style supplement that regularly included at least two lengthy joke columns. In short, it is clear that jokes were a pervasive and popular aspect of Victorian culture and that Victorian society was a good deal more amused than the old stereotype would tend to suggest. The fact that they valued jokes so highly and produced them in such remarkable quantity is enough justification for historians to take them seriously.

But what else do these jests offer to modern audiences? It is fair to say that many of them have not aged well. Take the following examples from Hugh Rowley's *Puniana* (1867) — a 250-page book of puns that is available to read on the *Internet Archive*:

What's the difference between a mouse and a young lady?  
One wishes to harm the cheese, the other to charm the he's.<sup>12</sup>

Why, when a very fat man gets squeezed coming out of the opera, does it make him complimentary to the ladies?  
Because the pressure makes him flatter. (p. 166)

If you were going to kill a conversational goose (how many of them do we not know!) what vegetable would she allude to?  
Ah!-spare-a-geese! (p. 214)

Why is Lord Overstone like a Britannia-metal teapot?  
Because he's a-Lloyd with lots of tin! (p. 157)

Perhaps surprisingly, these jokes were popular enough with Victorian readers to inspire a sequel, *More Puniana*, in 1875.<sup>13</sup> Modern audiences will probably be less enthusiastic. Some of the puns are still intelligible and might produce the odd smile, but others have become detached from the social, cultural, and political contexts that once underpinned their humour. The joke about Lord Overstone, for example, requires us to be aware that he

<sup>12</sup> Hugh Rowley, *Puniana; or, Thoughts Wise, and Other-Wise* (London: Hotten, 1867), p. 163 <<https://archive.org/details/punianaoorowlgoog>> [accessed 5 October 2015].

<sup>13</sup> *More Puniana; or, Thoughts Wise and Other-Why's*, ed. by Hugh Rowley (London: Chatto and Windus, 1875) <<https://archive.org/details/morepunianaoorowlgoog>> [accessed 5 October 2015].

was a banker ('lots of tin', i.e. money) with the surname Lloyd, and also to pick up on the fact that Britannia-metal teapots were alloyed ('a-Lloyd') with tin. In 1867, Rowley evidently expected his readers to possess this contextual knowledge to be able to recognize these seemingly obscure reference points and decode them quickly. A century and a half later, most readers will need to conduct background research in order to 'get' the joke.

But the very thing that kills such jokes for present-day audiences makes them all the more fascinating to historians. As Jan Bremmer and Herman Roodenburg argue, humour can offer a powerful 'key to the cultural codes and sensibilities of the past'.<sup>14</sup> Jokes, in particular, often work by subverting or recognizing the pre-existing expectations of their audience; they find humour by playing with an idea, person, situation, or piece of information which an audience, in a particular time and place, is expected to recognize. As a result, even the briefest of one-liners are often encoded with the attitudes, knowledge, and experiences of their intended audiences. By decoding a joke's reference points and unpacking the dynamics of its humour, historians can begin to access the minds of its readers in new ways. Historians of the early modern period have already begun to use the period's jestbooks and popular prints to shed new light on the everyday workings of a historically distant society.<sup>15</sup> In particular, Gatrell has demonstrated how a study of eighteenth-century popular humour 'can take us to the heart of a generation's shifting attitudes, sensibilities and anxieties' (p. 5). In particular, these sources allowed Gatrell to challenge dominant conceptions of the late-Georgian period as an age of politeness. The humour of the period and the way it was consumed, he argues, offers compelling evidence for an extensive metropolitan culture of *impoliteness*. In other words, the study of jokes and other forms of popular humour allowed him to say new and important things about the social and cultural values of an age; it is, he rightly argues, 'as plausibly a historian's subject as any other' (Gatrell, p. 5).

What happens when we subject Victorian jokes to similar forms of in-depth analysis? The gag about Lord Overstone seems unpromising at first glance, but it reveals something about the banker's fame and the pervasive nature of financial institutions in Victorian society. Readers of the joke were expected to know his name, his profession, and wealth, and to recognize him as a plausible reference point for jokes about money. Of course, we should not place too much interpretive weight on a single joke published in a relatively obscure book. But a wider search of books and newspapers reveals a broader trend. A near-identical joke was reportedly

<sup>14</sup> *A Cultural History of Humour: From Antiquity to the Present Day*, ed. by Jan Bremmer and Herman Roodenburg (Cambridge: Polity Press, 1997), p. xi.

<sup>15</sup> Simon Dickie, 'Hilarity and Pitilessness in the Mid-Eighteenth Century: English Jestbook Humour', *Eighteenth-Century Studies*, 37 (2003), 1–22; and Ian Munro, 'A woman's answer is neuer to seke': *Early Modern Jestbooks* (Aldershot: Ashgate, 2007).

told by the writer Shirley Brooks at a dinner of *Punch* contributors in March 1859 (apparently Thackeray laughed ‘heartily’), and appeared in numerous newspapers throughout the country.<sup>16</sup> Most of them reprinted it word for word, though the London correspondent of the *Belfast News-Letter* decided to help his readers decipher the punchline by adding Overstone’s surname and explaining the ‘alloy/a-Lloyd’ pun.<sup>17</sup> In 1863, it surfaced again in the *Sherborne Mercury* and an anthology of *Riddles and Jokes* compiled by the editor of *Every Boy’s Magazine*.<sup>18</sup> Rowley’s retelling of the joke was published in *Puniana* in 1867, which might then have inspired a reprint in the *Bedfordshire Times and Independent*.<sup>19</sup> It was revived again three years later by *Judy* magazine and reprinted in several other newspapers, including the popular *Lloyd’s Weekly*.<sup>20</sup> As I have argued elsewhere, the reprinting of jokes often transcended national boundaries, and this case study was no exception.<sup>21</sup> By the end of 1870, the joke about Overstone had reached the *Huron Expositor* (Ontario, Canada) and the *Hamilton Spectator* (Hamilton, Australia).<sup>22</sup> In 1885, two years after Overstone’s death, it was still circulating in the ‘Conundrums’ column of the *Australian Town and Country Journal*.<sup>23</sup> In other words, it appears that Lord Overstone remained a topical reference point for more than a decade, and one that was not just familiar to London’s literary and political circles but to the large working-class audience of *Lloyd’s Weekly*, along with readers in Lancaster, Portsmouth, Belfast, Cardiff, Canada, and Australia.

<sup>16</sup> M. H. Spielmann, *The History of Punch* (London: Cassell, 1895), p. 69; ‘FACETLÆ’, *North Wales Chronicle*, 2 April 1859, p. 3; ‘Wit and Humour’, *Hampshire Advertiser*, 2 April 1859, p. 3; ‘Random Readings’, *Family Herald*, 16 (2 April 1859), p. 784; ‘Odd Bits’, *Leeds Times*, 9 April 1859, p. 3; ‘SCRAPS’, *Manchester Times*, 16 April 1859, p. 3; ‘Varieties’, *Cheshire Observer*, 16 April 1859, p. 8; ‘Varieties’, *Leeds Intelligencer*, 30 April 1859, p. 3; ‘Varieties’, *Berrow’s Worcester Journal*, 7 May 1859, p. 3; [untitled], *Westmorland Gazette*, 7 May 1859, p. 4; ‘Literature, Science, and Art’, *Berkshire Chronicle*, 7 May 1859, p. 7; and ‘Varieties’, *Hereford Times*, 21 May 1859, p. 6.

<sup>17</sup> ‘London Correspondence’, *Belfast News-Letter*, 26 April 1859, p. 3.

<sup>18</sup> ‘Our Christmas Riddles’, *Sherborne Mercury*, 22 December 1863, p. 4; *Riddles and Jokes Collected by the Editor of Every Boy’s Magazine* (London: Routledge, Warne, and Routledge, 1863), p. 43.

<sup>19</sup> ‘Our Arm Chair’, *Bedfordshire Times and Independent*, 17 August 1867, p. 2.

<sup>20</sup> ‘Extracts from the Comic Papers’, *Belfast News-Letter*, 28 May 1870, p. 4; ‘Extracts from the Comic Papers’, *Lancaster Gazette*, 28 May 1870, p. 2; ‘Cuttings from the Comic Journals’, *Lloyd’s Weekly Newspaper*, 29 May 1870, p. 2; ‘Wit and Humour’, *Cardiff and Merthyr Guardian*, 4 June 1870, p. 3; and ‘Varieties’, *Jackson’s Oxford Journal*, 18 June 1870, p. 6.

<sup>21</sup> Bob Nicholson, “‘You Kick the Bucket; We Do the Rest!’: Jokes and the Culture of Reprinting in the Transatlantic Press”, *Journal of Victorian Culture*, 17 (2012), 273–86.

<sup>22</sup> ‘Varieties’, *Huron Expositor*, 9 September 1870, p. 1; and ‘Two Charming Little Experiments’, *Hamilton Spectator*, 20 August 1870, p. 5.

<sup>23</sup> ‘Conundrums’, *Australian Town and Country Journal*, 3 January 1885, p. 31.

This is just one small example. There are thousands — possibly even millions — of other Victorian jokes waiting to be discovered and unpacked. Tracing their recurring themes, tropes, and reference points promises to reveal valuable new insights into nineteenth-century culture and society. By examining the things that made ordinary Victorians laugh, we can explore the shared beliefs and practices that underpinned a range of social, cultural, political, and economic relations. A typical newspaper's joke column might shed light on: the gender relations in Victorian marriages; public perceptions of tramps and lawyers; discourses of race and class; the perceived shortcomings of the French; the cultural significance of local dialects; and a range of other characters, situations, and attitudes that millions of readers were expected to recognize in order to 'get' the joke. Crucially, jokes help to reconstruct the attitudes and experiences of communities who are often under-represented in the historical record. For example, I initially became interested in newspaper jokes when using them as a way to challenge prevailing academic assumptions about Victorian attitudes to America, which tend to be over-reliant on the work of middle-class travel writers.<sup>24</sup> Finally, the periodicity of newspapers allows us to see how ideas and practices captured in these jokes changed over time. How, we might wonder, did Victorian joke writers respond to changes in gender relations and the emergence of the women's suffrage movement? How did jokes about Charles Darwin and his scientific theories develop over the course of the century? And what do jokes tell us about the racial theories that underpinned Britain's imperial activities? In short, there is much to be gained from unearthing these forgotten fragments of Victorian humour and subjecting them to sustained historical analysis.

### Remixing the archive

Finding and exploring Victorian jokes poses a number of practical and methodological challenges. If researchers want to read a random sample of the period's humour, then it is fairly straightforward to access dedicated joke books such as *The London Budget of Wit* (1817), *English Jestes and Anecdotes* (1880), *The American Joe Miller* (1865), or Rowley's *Puniana* (1867).<sup>25</sup> But, as we have seen, these dedicated anthologies represent just a fraction of the surviving record of nineteenth-century jests; many more were scattered

<sup>24</sup> Bob Nicholson, 'Jonathan's Jokes: American Humour in the Late-Victorian Press', *Media History*, 18 (2012), 33–49; and Bob Nicholson, 'Looming Large: America and the Late-Victorian Press, 1865–1901' (unpublished doctoral thesis, University of Manchester, 2012).

<sup>25</sup> *The London Budget of Wit; or, A Thousand Notable Jestes* (London: Walker and Edwards, 1817) <<https://archive.org/details/londonbudgetwitoomirtgoog>>; *English Jestes and Anecdotes Collected from Various Sources* (Edinburgh: Paterson, 1880) <<https://archive.org/details/englishjestsanecooedin>>; and *The American Joe Miller: A Collection of Yankee Wit and Humour*, comp. by Robert Kempt, 2nd edn (London: Adams and Francis, 1865) <<https://archive.org/details/americanjoemilllookempgoog>> [all accessed 5 October 2015].

throughout newspapers and magazines. Finding these jokes is more problematic. Until relatively recently, historians would have had little choice but to leaf through millions of pages of print in search of anything with a punchline, and locating a joke on a specific topic would have been simply impracticable. Fortunately, the digitization of nineteenth-century books and newspapers has disrupted conventional ‘top-down’ approaches to browsing the archive and opened up the possibility of ‘bottom-up’ keyword searching for specific words and phrases.<sup>26</sup> Under the old system of manual browsing, it would have taken years of reading (and a few long-haul flights to Canada and Australia) to track down reprints of the Overstone joke. Now, we can simply enter a phrase from the joke (i.e. ‘why is Lord Overstone’) into a series of online archives and analyse the results.<sup>27</sup> It is important to remember that these digital searches are by no means comprehensive. Most nineteenth-century print culture remains un-digitized, and existing archives are undermined by the variable quality of their OCR data.<sup>28</sup> It is likely that the Overstone joke appeared in several other publications; in particular, I suspect that it was published in a prominent newspaper or magazine in 1859 before suddenly ‘going viral’ throughout the provincial press. Nevertheless, despite these blind spots, it is clear that keyword searching represents a huge step forward for this kind of cultural analysis.

However, the methodology used to trace the Overstone joke has undeniable limitations. I was only able to locate its reprints because I had

<sup>26</sup> For a discussion of these ‘top-down’ and ‘bottom-up’ approaches to reading newspapers, see Bob Nicholson, ‘The Digital Turn: Exploring the Methodological Possibilities of Digital Newspaper Archives’, *Media History*, 19 (2013), 59–73.

<sup>27</sup> In this instance, I used:

- a) Gale Cengage’s subscription-based *NewsVault* platform to search across multiple archives of British newspapers and periodicals.
- b) ProQuest’s subscription-based *British Periodicals* to search for the joke in weightier periodicals.
- c) D. C. Thomson’s subscription-based *British Newspaper Archive* to search a wider range of local newspapers.
- d) The newspaper search engine *Elephind* to search a range of archives in Australia and North America.
- e) The *Internet Archive* and *HathiTrust* to find mentions of the joke in books.
- f) A simple Google search to try and mop up any other hits — this also searches the *Internet Archive* and Google’s own archive of historic newspapers.

All the above websites accessed 5 October 2015.

<sup>28</sup> Simon Tanner, Trevor Muñoz, and Pich Hemy Ros, ‘Measuring Mass Text Digitization Quality and Usefulness: Lessons Learned from Assessing the OCR Accuracy of the British Library’s 19th Century Online Newspaper Archive’, *D-Lib Magazine*, 15.7/8 (July/August 2009) <<http://www.dlib.org/dlib/july09/munoz/o7munoz.html>> [accessed 5 October 2015].

Rowley's version of the joke. I selected a relatively distinctive string of words from the text ('why is Lord Overstone') and used this to focus my search. But imagine if we did not have this starting point and wanted to find out if any jokes were written about Overstone during the nineteenth century. Where would we begin? A full-text search for 'Overstone' in the *19th Century British Library Newspapers* database returns 6892 results — the vast majority of which are not jokes. In information retrieval theory, the relevance of a search is measured by its 'precision' and 'recall', and the precision of a search is determined by the percentage of results that are relevant to the task in hand. In this case, we could try to improve the precision by limiting it to articles with title words such as 'wit', 'humour', 'jokes', 'jests', 'comic', 'varieties', and 'scraps', but these searches bypass many relevant sources (jokes were often published as untitled column fillers), and still require researchers to have a clear search term in mind. Imagine if we were not interested in Overstone himself, but in jokes about Victorian bankers — we might have selected search terms like 'bank', 'banker', 'banking', or 'money', but these would not have located the joke about Overstone. In information retrieval theory, these searches may be better in terms of their precision, but they are worse at 'recall' — this refers to the fraction of relevant results that are retrieved by the search. In this case, overly precise searches run the risk of missing lots of relevant results. This balance between precision and recall is a problem when attempting to retrieve any kind of joke from existing newspaper databases. Imagine, for instance, that we wanted to find Victorian jokes about politics, literature, sport, or marriage. How many potential searches would we need in order to locate all of the gags on such broad topics? How would we disentangle jokes about cricket from a newspaper's sporting columns, or jokes about divorce from its courtroom coverage? Even if we could focus purely on humour columns, how would we pick out jokes *about* marriage from the thousands of gags that feature words like 'wife' or 'husband' but are actually about something else?

With time and perseverance we might have limited success, but we would be forced to 'go against the grain' of an archive that was (understandably) never designed to support this kind of specific research. What is more, we would need to conduct our research across a range of different archives, each with their own search interfaces, organizational systems, and technical idiosyncrasies. Only the most enthusiastic joke-hunters are likely to persevere in such circumstances. In short, if we really want to unlock the potential of Victorian jokes and subject them to in-depth historical analysis, then we need to find a way to make them more accessible.

There are ways that this might be achieved using existing archives. The *British Newspaper Archive (BNA)* allows users to 'tag' articles with descriptive keywords, which would potentially allow us to curate a virtual

sub-collection of joke columns. This archive also allows users to correct errors in the OCR text, which would improve the quality of the collection and the discoverability of its jokes. However, it would require us to operate within the confines of a single archive — one that requires users to pay a subscription — and to ignore other important sources. Moreover, the structure of this database means that we could only organize our collection by column, rather than by individual jokes. This means that we could not tag individual gags with their genre, subject, author, or country of origin. As a result, the discovery of jokes would still be dependent on speculative keyword searches. Consider the following example:

#### A USEFUL PRESENT

- Mrs. Henry Peck (whose mother has been visiting them for over four months): ‘I don’t know what to buy mother for a Christmas present. Do you?’
- Mr. Henry Peck: ‘Yes! Buy her a travelling bag!’<sup>29</sup>

This is a standard mother-in-law joke, but the term ‘mother-in-law’ is never used. As a result, a keyword search for the phrase ‘mother-in-law’ will not find the joke. We might broaden our search to the word ‘mother’, but this will return thousands of jokes that are not specifically about mothers-in-law and will still ignore gags which use words like ‘mama’ instead of ‘mother’. Moreover, users would have no way to focus their research on a particular form of humour (e.g. puns, conversational jokes, comic songs, funny definitions, etc.). In short, working within the confines of an existing archive like the *BNA* would not allow us to catalogue jokes using our own categories of metadata — the most we could hope to do is highlight the *presence* of jokes in specific articles, and then leave researchers to do the rest. In other words, this would be a small and rather compromised step forward.

The alternative is to build a new archive dedicated to Victorian jokes. This might seem like a mammoth undertaking; anybody who has been involved in a digitization project will attest to how complicated and expensive these endeavours can be. In a time of shrinking budgets, I doubt that any funding bodies will be anxious to spend millions of pounds on a database of laboured Victorian puns. However, thousands of Victorian jokes have *already* been digitized — we simply need to find a way to extract them from existing databases. Then, we could recompile them into a new data set designed to meet the specific requirements of historical research. This would allow us to create a more focused research environment and to catalogue jokes according to genre, subject, author, date, characters, place of publication, or any other categories that scholars deem useful. The practical benefits of this approach are clear. But this process of combing and

<sup>29</sup> ‘American Jokes’, *Lloyd’s Weekly Newspaper*, 5 January 1896, p. 8.

remixing elements of existing digital collections has other advantages too. Firstly, it encourages us to disrupt the artificial boundaries created by long-established archives and to recognize the rich, intertextual connections that existed within Victorian culture. For example, we currently tend to divide the period's books, newspapers, and periodicals into separate archival and disciplinary silos; each of them are held in dedicated digital collections and are often studied by unconnected communities of scholars. And yet, when these texts were originally published, they sat side by side in nineteenth-century bookshops and reading rooms. Victorian readers moved effortlessly across the genre boundaries that now define our archives and shape the contours of research. Moreover, as the Overstone joke demonstrates, texts and ideas were equally mobile; they leapt from newspapers to books, then to the dinner table and back again with an agility that historians struggle to match. An archive dedicated to Victorian jokes would help to facilitate a more interdisciplinary and intertextual approach to the period.

Of course, the creation of an archive dedicated to jokes would establish a new and equally problematic set of structures, boundaries, and perspectives. In particular, by extracting jokes and isolating them we disconnect them from the contexts in which they were originally read and published. Can we really understand how the readers of *Lloyd's Weekly* deciphered the Overstone joke without also considering how he was represented within the paper's political coverage? Similarly, by treating newspaper gags as self-contained textual units, we might lose sight of how multi-joke columns were compiled, presented, and consumed. Readers of *Lloyd's* would have encountered fifteen other jokes in that week's column of 'Cuttings from the Comic Journals' before reaching the one about Overstone — what role did these gags play in shaping readers' responses to our joke? Moreover, might these readers have approached the Overstone joke in a different frame of mind when it was printed in a column entitled 'Riddles' rather than 'Comic Cuts'?

In truth, the isolation of information is already a problem with current digital archives. Historians are becoming increasingly blasé about decontextualized reading; search engines allow us to jump instantly to a word and sentence within an article without examining the material that surrounds it. These 'keyword blinkers' are a problem — one that a new joke archive might exacerbate.<sup>30</sup> As a result, it seems that we are caught between two extremes. If we continue to work within existing archives, then we must labour under the weight of an almost suffocating amount of context, an overwhelming sea of information that obscures our sources and masks the connections that existed between different forms of print. On the other hand, by extracting jokes from their original sources and presenting them

<sup>30</sup> Nicholson, 'The Digital Turn', p. 61. See also Adrian Bingham, 'The Digitization of Newspaper Archives: Opportunities and Challenges for Historians', *Twentieth Century British History*, 21 (2010), 225–31 (pp. 229–30).

in a decontextualized fashion, we undermine our ability to understand their historical workings and significance. Fortunately, there is a middle ground to be found. A dedicated archive would display the essential contextual information for each joke (original publication title, column title, date, page number, etc.), and therefore allow researchers to replace a gag within its original context.

For an existing example of this approach, consider the impressive *Reading Experience Database* (*RED*). This archive contains evidence of historical ‘reading experiences’ extracted from a broad range of sources. Here, for example, are the [results of a search for sources that mention readers of \*Lloyd’s Weekly Newspaper\*](#). They include material from Henry Mayhew’s *London Labour and the London Poor* (1861); an article from *The Times* (1845); M. V. Hughes’s autobiographical account of *A London Family: 1870–1900* (published 1946); and evidence from Jonathan Rose’s *Intellectual Life of the British Working Classes* (2001). Each of these sources would normally be organized into separate archives and libraries, but the *RED* database brings them together in a new way. Crucially, it does not contain these sources in full, but provides short extracts that relate to the reading of *Lloyd’s Weekly*. For example, the entry from Hughes’s memoir simply reads: ‘How horrified my father was on discovering that the servants had been reading little bits to me out of “Lloyd’s Weekly” [on a Sunday].’<sup>31</sup> This extract is supported by a remarkable amount of metadata. We are presented with information about the reader/listener of the newspaper, including her name, age, gender, nationality, and socio-economic class. We are also informed about the title, genre, and provenance of the text being read, along with details on the place and context in which it was read. This rich metadata allows users to construct sophisticated searches that move beyond the use of keywords. For example, we can immediately identify all entries in the database that feature adult female servants reading newspapers. This kind of search could not be accomplished using keywords alone. Crucially, each entry in the *RED* database also contains full bibliographical details for the original source, which makes it easy for researchers to reunite extracts with their original context. A dedicated Victorian jokes database would work in a similar way. Each joke would have a rich set of descriptive metadata, including all of the contextual information required to connect the joke with its original place of publication. Ideally, we would make this process easier by including a copy of, or direct link to, the original source — though the viability of this feature would depend on a range of legal and technical considerations.

---

<sup>31</sup> M. V. Hughes, *A London Family: 1870–1900* (London: Oxford University Press, 1946), p. 73 <[http://www.open.ac.uk/Arts/reading/UK/record\\_details.php?id=906](http://www.open.ac.uk/Arts/reading/UK/record_details.php?id=906)> [accessed 5 October 2015].

Finally, a dedicated Victorian jokes database — or indeed any similar digital project — should be built around the principles of open access.<sup>32</sup> This would pose the perennial problems of funding and sustainability, but the cost of running a project like this should not be too onerous — after all, it is unlikely to receive millions of visitors. An open approach has two key advantages. Firstly, it ensures that researchers and members of the public will be able to access the collection without paying a subscription. As we will see in the next section, the participation of volunteers will be vital to the success of the project. Secondly, the creation of an open and carefully curated data set will enable researchers to develop their own approaches to the archive. For example, scholars could download the data set and use other pieces of software to map the circulation of jokes using GIS; track the rise and fall of words and phrases using tools like Google’s ngram viewer; quantify the changing popularity of particular joke genres, subjects, and characters; search for reprinted jokes using plagiarism detection tools; or unpack linguistic patterns using corpus analysis software. The award-winning *Old Bailey Online* project has already demonstrated the value of opening archives up to these acts of ‘remixing’. In its original format, the project allowed users to browse and search trial records using a fairly conventional archival interface. However, the data set has now been used for a variety of other projects. For example, *Locating London’s Past* allows users to search *Old Bailey Online* and map their results onto a historical map of London. Similarly, the *Data Mining with Criminal Intent* project has developed a system for exporting the archive’s trial records to Voyant Tools — a suite of free, online text analysis software that allows users to produce word clouds, concordances, and other visualizations. These tools allow us to ‘distant read’ thousands of historical trial records and discover new underlying themes and patterns in ways that were not possible using the basic search interface. Finally, *Old Bailey Online* is also searchable via the *Connected Histories* platform. This offers a new way to access the archive, one that helps researchers to trace references to people and places across a range of other archives and databases. All of this was made possible by the openness of *Old Bailey Online* and the ease with which users can access, download, and repurpose its data.

---

<sup>32</sup> For reflections on the challenges and benefits of open access, see Peter Suber, *Open Access* (Cambridge, MA: MIT Press, 2012) <[http://mitpress.mit.edu/sites/default/files/titles/content/9780262517638\\_Open\\_Access\\_PDF\\_Version.pdf](http://mitpress.mit.edu/sites/default/files/titles/content/9780262517638_Open_Access_PDF_Version.pdf)> [accessed 5 October 2015]; Martin Paul Eve, *Open Access and the Humanities: Contexts, Controversies and the Future* (Cambridge: Cambridge University Press, 2014) <<http://dx.doi.org/10.1017/CBO9781316161012>>; Jerome McGann, *A New Republic of Letters* (Cambridge, MA: Harvard University Press, 2014); M. H. Beals, ‘Rapunzel and the Ivory Tower: How Open Access Will Save the Humanities (from Themselves)’, *Journal of Victorian Culture*, 18 (2013), 543–50.

Unfortunately, not all digitization projects have matched the gold standard set by *Old Bailey Online*. A lack of public funding — and, in some cases, a lack of imagination — has led to the creation of archives that are encumbered by paywalls and complex copyright restrictions, or structured in ways that makes data difficult to extract and reuse. Victorian newspapers and periodicals have been digitized in a particularly restrictive and commercially minded fashion, especially when compared to the openness of similar archives in North America, Australia, New Zealand, Wales, and Europe.<sup>33</sup> But it is not too late to begin liberating them. Recent changes in British copyright law have reinforced our right to directly access data sets from commercial archives and explore them using our own digital tools.<sup>34</sup> The *Distant Reading Early Modernity (DREaM)* project provides a recent example of how old data sets are now being used in this fashion. This endeavour centres on the creation of a corpus of 44,000 early modern texts using data from ProQuest's *Early English Books Online (EEBO)* database. As they put it:

Unlike EEBO, DREaM will enable mass downloading of custom-defined subsets rather than obliging users to download individual texts one-by-one. In other words, we have designed it to function at the level of 'sets of texts,' rather than 'individual texts.' Examples of subsets one might potentially generate include 'all texts by Ben Jonson,' 'all texts published in 1623,' or 'all texts printed by John Wolfe.' [...] The ability to generate custom-defined subsets is important because it will allow researchers to interrogate the early modern canon with distant reading tools such as David Newman's Topic Modeling Tool or the suite of visualization tools available on Voyant-tools.org. By paving the way for these possibilities, DREaM will significantly expand the range and sophistication of technologies

<sup>33</sup> For open access newspaper archives see:

- a) United States: *Chronicling America*.
- b) Australia: *Trove*.
- c) New Zealand: *Papers Past*.
- d) Wales: *Welsh Newspapers Online*.
- e) Europe: *Europeana Newspapers*. This collection is also accessible through the main *Europeana* portal.
- f) A fairly comprehensive list of online newspapers is also maintained on Wikipedia: see <[https://en.wikipedia.org/wiki/Wikipedia:List\\_of\\_online\\_newspaper\\_archives](https://en.wikipedia.org/wiki/Wikipedia:List_of_online_newspaper_archives)>.

All the above archives were accessed 5 October 2015.

<sup>34</sup> 'Exceptions to Copyright: Research' (Newport: Intellectual Property Office, 2014) <<https://www.gov.uk/guidance/exceptions-to-copyright>> [accessed 6 October 2015].

currently available to researchers who wish to gain a broad sense of printed matter in the period.<sup>35</sup>

This is a good example of how a long-established archive can be remixed in ways that encourage new forms of research. An open data set of Victorian jokes, extracted from a range of existing digital archives, would open up similarly new methodological possibilities. All we need to do now is work out how to build it.

### The Victorian Meme Machine

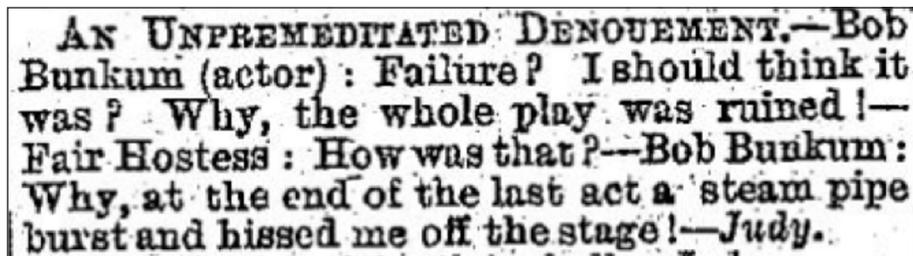
The extraction of jokes from existing digital archives turns out to be a complex process. Each archive and source type requires a different approach. As a starting point, the British Library Labs team and I decided to focus our attention on two of the Library's digital collections: the *19th Century British Library Newspapers* database, and a collection of nineteenth-century books digitized by Microsoft in 2008.

While it is possible to find jokes elsewhere, these sources provide the largest concentrations of material. Dedicated joke books, such as *Puniana*, contain hundreds of jokes in a single package. These are the easiest texts to process, because their jokes are not mixed in with other material and the quality of their OCR data is usually good. For the pilot stage of the project, one volunteer, Wendy Durham, copied 728 jokes from the OCR transcription of the *Book of Humour, Wit and Wisdom* (1880) and pasted them into an Excel spreadsheet that could be imported into our database. In the fullness of time it will be necessary to automate this process, but a manual approach gave us a quick and useful sample of jokes to develop.

Working with the newspaper archive is more complex. Some newspapers mixed humour with other pieces of entertaining miscellany under titles such as 'Varieties' or 'Our Carpet Bag'. The same is true of books, which often combined jokes with short stories, comic songs, and material for parlour games. While it is easy to find these collections, recognizing jokes is more problematic. As our project develops, we intend to experiment with natural language processing to build a joke-detection tool that can pick out new content with similar formatting and linguistic characteristics to jokes that we have already found. For example, conversational jokes usually have capitalized names (or pronouns) followed by a colon and, in some cases, include a descriptive phrase enclosed in brackets. So, if a text includes strings of characters like 'Jack (...):' or 'She (...):' then there is a good chance that it might be a joke. Similarly, many jokes begin with a

<sup>35</sup> Stephen Wittek, 'Introducing DREaM', *Early Modern Conversations* <<http://earlymodernconversations.com/introducing-dream/>> [accessed 5 October 2015].

capitalized title followed by a full stop and a hyphen, and end with an italicized attribution. Here is a characteristic example of all three types (*Fig. 1*):



*Fig. 1:* A typical Victorian newspaper joke. ‘Jokes of the Day’, *Lloyd’s Weekly Newspaper*, 5 April 1891, p. 7.

Unfortunately, conventional search interfaces are not designed to recognize nuances in punctuation, so we will need to build something new ourselves. However, for the pilot stage of the project, we chose to focus our efforts on harvesting the ‘low-hanging fruit’ found in clearly defined joke columns.

In order to locate these columns we have compiled a continually expanding list of search terms. Obvious keywords like ‘jokes’ and ‘jests’ are most effective, but we have also found material using other words like ‘quips’, ‘cranks’, ‘wit’, ‘fun’, ‘jingles’, ‘humour’, ‘laugh’, ‘comic’, ‘snaps’, and ‘siftings’. Yet while these general search terms are useful, they do not catch everything. Consider these peculiarly named columns from the *Hampshire Telegraph* (*Fig. 2*). At first glance, they look like food recipes, but in point of fact they are columns of imported American jokes named after a popular Yankee delicacy. Uncovering material like this is more laborious and requires manual searches for peculiarly named books and joke columns.

In the case of newspapers, this approach requires some educated guesswork. Most joke columns appeared in popular weekly papers, or in the weekend editions of mass-market dailies. Similarly, while the placement of jokes varied from paper to paper (and sometimes from issue to issue), they were typically located at the back of the paper alongside children’s columns, fashion advice, recipes, and other miscellaneous titbits of entertainment. Finally, once a newspaper has been found to contain one set of joke columns, it is likely that more will be found under other names. For example, the *Newcastle Weekly Courant* discontinued its long-running ‘American Humour’ column in 1888 and renamed it ‘Yankee Snacks’. Tracking a single change of identity like this is fairly straightforward; once the new title has been identified we simply add it to our list of search terms. Unfortunately, the editorial whims of some newspapers are harder to follow. For example, the *Hampshire Telegraph* often scattered multiple joke columns throughout a single issue. To make things more complicated, they tended to rename

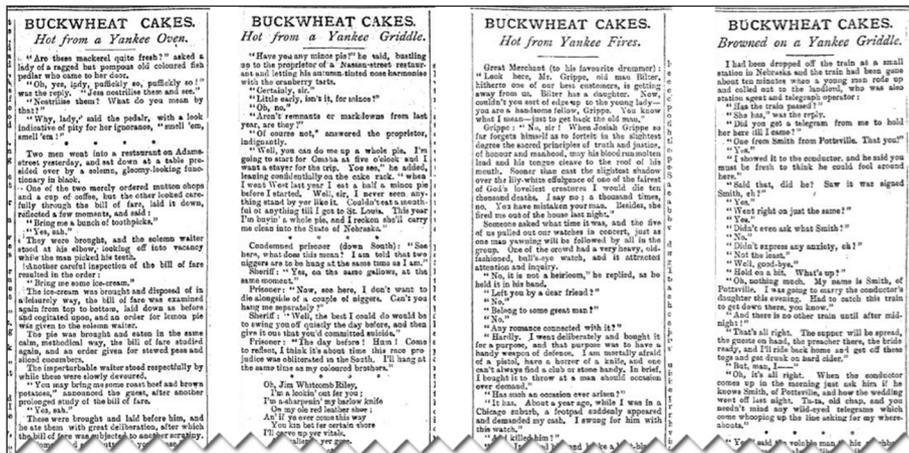


Fig. 2: Columns of American humour published in the *Hampshire Telegraph*, 14 February 1891 to 4 April 1891.

and reposition these columns every two weeks. Here is a sample of the paper's American humour columns, all drawn from the first six months of 1892 (Fig. 3). For papers like this, the only option at the moment is to manually locate joke columns one at a time. In other words, while our initial set of keywords should enable us to find and extract thousands of joke columns relatively quickly, more nuanced (and more laborious) methods will be required to get the rest.

Using a range of keywords and manual browsing methods we soon managed to identify the locations of at least one hundred thousand newspaper jokes. In truth, this was always going to be the easy task. The real

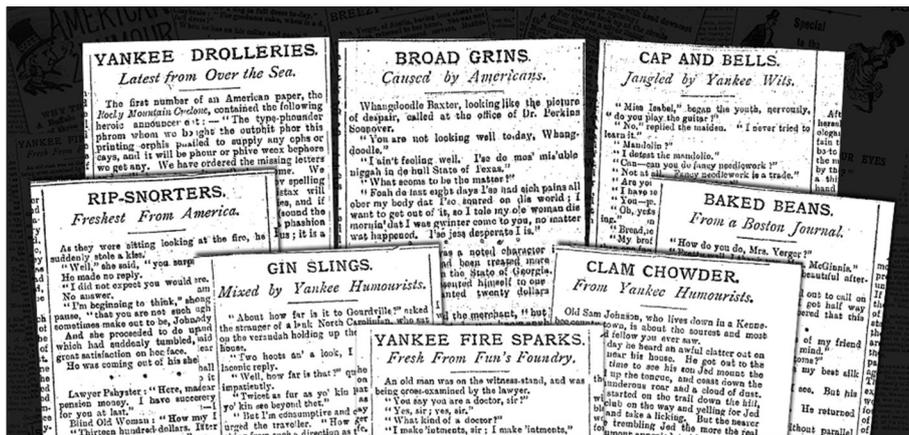
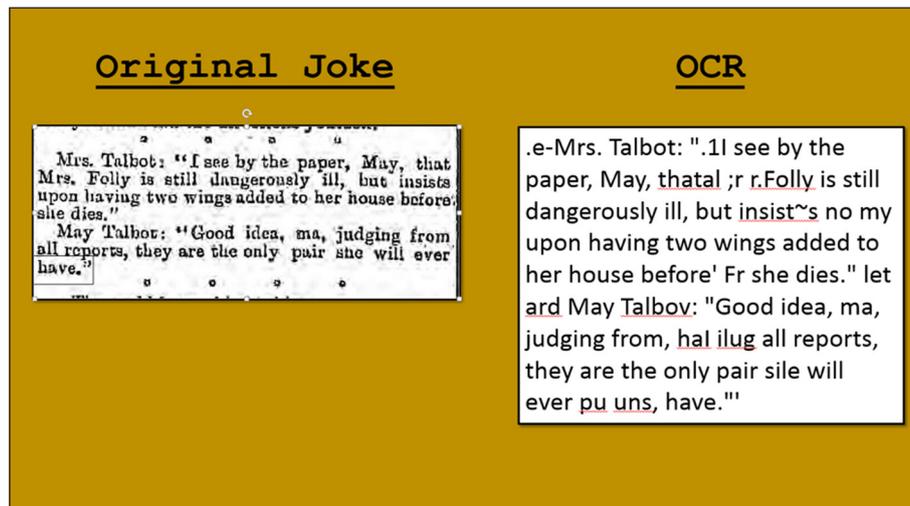


Fig. 3: Columns of American humour published in the *Hampshire Telegraph* in 1892.

challenge lies in extracting these jokes from their archives, importing them into our database, and then converting them into a format that can be searched by researchers and broadcast over social media. On the plus side, the underlying structure of the *19th Century British Library Newspapers* database is fairly well suited to our purposes. Newspaper pages have already been broken up into individual articles and columns. As a result, it should theoretically be possible to isolate every article with the title 'Jokes of the Day' and then extract them from the rest of the database. When I suggested this project to the BL Labs, I naively thought we could perform these extractions in a matter of minutes — unfortunately it is not that easy. The archive's public-facing platform is owned and operated by the commercial publisher Gale Cengage, who sell subscriptions to universities and libraries around the world (UK universities currently get free access via JISC). Consequently, access to the archive's underlying content is restricted when using this interface. It does not provide access to the underlying data set, and images can only be downloaded one by one using a web browser's 'save image as' button. In other words, we cannot use the commercial interface to instantly grab the files for every article with the phrase 'Jokes of the Day' in its title. The British Library keeps its own copy of these files, but when our project began they were stored in a format that was difficult to access and extremely cumbersome to search. In order to move forward with the automatic extraction of jokes we needed to secure access to this data, transfer it onto a more accessible storage system, build an index that allows us to search for column titles, find a way to simultaneously extract all of the relevant XML files and their associated TIFFs, and then finally import them into a new database. All of this is technically possible and the Labs team are currently working to make the data set more accessible to all researchers who want to pursue work of this nature. However, the first stage of the Victorian Meme Machine project was only funded for six months, so we decided to press ahead with a small sample of manually extracted columns and focus our attention on the next stages of the project.

For our sample we manually downloaded all of the 'Jokes of the Day' columns published by *Lloyd's Weekly News* in 1891. These columns contain a mixture of joke formats — puns, conversations, comic stories, etc. — and are formatted in a way that makes them broadly representative of the material found elsewhere in the database. If we can find a way to process 1000 jokes from this source, we should be able to scale things up to manage 100,000 similar jokes from other newspapers. Our sample of joke columns was downloaded as a set of JPG images. In order to make them keyword searchable and send them out in 'remixed' formats over social media we first needed to convert them into accurate, machine-readable text. Unfortunately, the quality of the original OCR makes it difficult to

fully automate this process. Here is an example of how one joke has been interpreted by OCR software (*Fig. 4*):



*Fig. 4:* The accuracy of OCR software. The joke comes from ‘Laugh and Grow Fat’, *Hampshire Telegraph*, 20 April 1889, p. 11.

Some jokes have been captured more successfully than this, but many are substantially worse. Joke columns often appeared at the edge of a page, which makes them susceptible to fading and page bending. They also make use of unusual punctuation, which tends to confuse OCR software. Unlike newspaper archives, which retain a limited degree of search-related functionality even with relatively low-quality OCR, our project requires 100 per cent accuracy (or something close) in order to remix the jokes and republish them in new formats.

For the pilot stage of the project, we built a temporary transcription platform using the open source publishing platform [Omeka](#), and a plug-in called [Scripto](#). It is not ideally suited to the specific needs of the project and will need to be replaced, but it enabled us to quickly set up a successful workflow without building a bespoke transcription platform. Here is an example of a typical transcription page (*Fig. 5*). This transcription process is currently open to a small group of volunteers: the project team, a couple of other members of staff at the British Library, and ten of my students at Edge Hill University. In the space of a few days, they managed to add 700 jokes to the database. This is a promising start and has provided us with some useful data, but if we want to create an archive of one million Victorian jokes then we will need to scale things up.

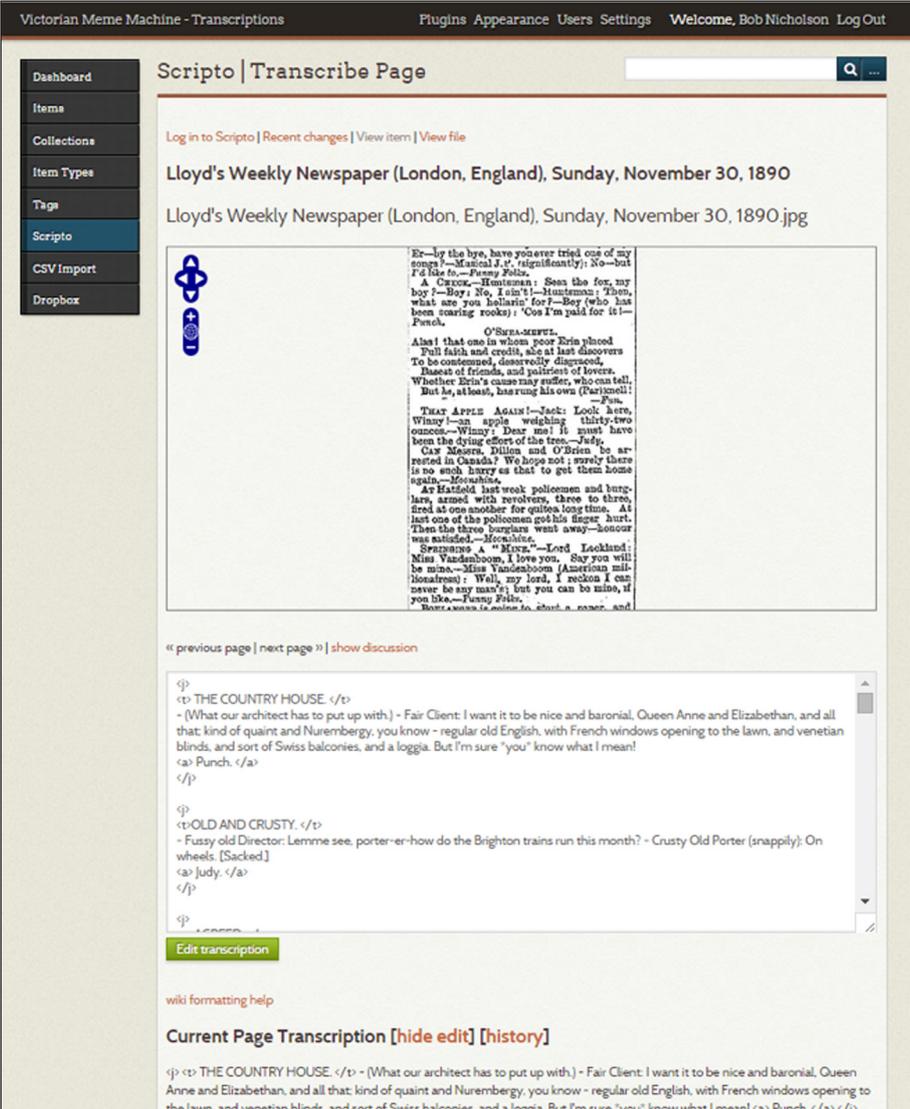
Victorian Meme Machine - Transcriptions Plugins Appearance Users Settings Welcome, Bob Nicholson Log Out

Scripto | Transcribe Page

Log in to Scripto | Recent changes | View item | View file

Lloyd's Weekly Newspaper (London, England), Sunday, November 30, 1890

Lloyd's Weekly Newspaper (London, England), Sunday, November 30, 1890.jpg



Er—by the bye, have you ever tried one of my songs?—'Madical J.', significantly; No—but I'd like to.—'Fenny Fells.

A. CROOK.—'Huntamen! Seen the fer, my boy?—'Boy: No, I ain't!—'Huntamen! Then, what are you hellarin' for?—'Boy (who has been roaring rooks): 'Ovs I'm yald for it!—'Punch.

O'SHEA-SHEFF.

Alas! that one in whose poor Erin's blood  
Full faith and credit, she at last discovers  
To be condemned, deservedly disposed,  
Basest of friends, and purest of lovers.  
Whether Erin's cause may suffer, who can tell,  
But he, at least, bearing his own (Partison):  
—'Pun.

THAT APPLE AGAIN!—'Jack: Look here,  
Wimpy I—can apple weighing thirty-two  
ounces.—'Wimpy: Dear me! It must have  
been the dying effort of the tree.—'Judy,  
Cox Moore, Dillon, and O'Brien be ar-  
rested in Canada? We hope not; surely there  
is no such hurry as that to get them home  
again.—'Moosehide.

At Hatfield last week policemen and bung-  
lars, armed with revolvers, three to three,  
fired at one another for quite a long time. At  
last one of the policemen got his finger hurt.  
Then the three burglars went away—honour  
was satisfied.—'Moosehide.

Serjeant A "Kite."—'Lord Locksted:  
Miss Vandebloom, I love you. Stay you will  
be mine.—'Miss Vandebloom (American mil-  
lionaire): Well, my lord, I reckon I can  
never be any man's; but you can be mine, if  
you like.—'Fenny Fells.

BOYLE AND HIS COLINE TO, about a corner, and

« previous page | next page » | show discussion

THE COUNTRY HOUSE. </p>  
<p>- (What our architect has to put up with.) - Fair Client: I want it to be nice and baronial, Queen Anne and Elizabethan, and all that kind of quaint and Nuremberg, you know - regular old English, with French windows opening to the lawn, and venetian blinds, and sort of Swiss balconies, and a loggia. But I'm sure "you" know what I mean! <a> Punch. </a>  
</p>

OLD AND CRUSTY. </p>  
<p>- Fuzzy old Director: Lemme see, porter-er-how do the Brighton trains run this month? - Crusty Old Porter (snappily): On wheels. [Sacked.] <a> Judy. </a>  
</p>

Edit transcription

wiki formatting help

Current Page Transcription [hide edit] [history]

<p>THE COUNTRY HOUSE. </p> - (What our architect has to put up with.) - Fair Client: I want it to be nice and baronial, Queen Anne and Elizabethan, and all that kind of quaint and Nuremberg, you know - regular old English, with French windows opening to the lawn, and venetian blinds, and sort of Swiss balconies, and a loggia. But I'm sure "you" know what I mean! <a> Punch. </a> </p>

Fig. 5: Transcribing Victorian jokes using Omeka and Scripto.

In recent years, crowdsourcing initiatives have become increasingly central to the development of new digital resources. Projects such as *Dickens Journals Online*, *Transcribe Bentham*, and the *Zooniverse*, not to mention the long-running *Project Gutenberg*, have all demonstrated what can be achieved by attracting an active community of volunteers to help tag and transcribe data. However, crowdsourcing projects face at least four key challenges: attracting a community of appropriate volunteers; keeping

this community of volunteers motivated; ensuring that volunteers produce high-quality work; and making sure that the free labour provided by these volunteers is organized in an ethical fashion. In order to attract and motivate volunteers, some digitization projects have attempted to borrow techniques developed in the computer game industry. As Scott Nicholson puts it:

A common implementation of gamification is to take the scoring elements of video games, such as points, levels, and achievements, and apply them to a work or educational context. While the term is relatively new, the concept has been around for some time through loyalty systems like frequent flyer miles, green stamps, and library summer reading programs. These gamification programs can increase the use of a service and change behavior, as users work toward meeting these goals to reach external rewards.<sup>36</sup>

For example, UCL's *Transcribe Bentham* project motivates its volunteers by displaying a regularly updated 'Top Contributors' league table. Users earn points by completing more transcriptions and have the ability to work up from the rank of 'Novice' (2500 points) to 'Prodigy' (75,000). Points are also earned by inviting friends to do transcriptions and by participating in online discussions. The project has now transcribed more than seventeen thousand handwritten manuscripts, which suggests the approach has merits, although it is possible of course that volunteers were more motivated by the opportunity to contribute to an important historical project than to rise up a league table.

The problem with current approaches to gamification is that they usually rely on an arbitrary and essentially meaningless scoring system of points and badges to reward the transcription effort. It appears that these techniques only motivate users for a relatively short time. When Chris Lintott, founder of the Zooniverse project, experimented with this kind of gamification he found that

the best people were systematically leaving, because once you switch into collecting points, and you find you're winning the game, then you get the impression that you've finished — mastered it — and you put the game down. We'd essentially built a system that drove away our best people.<sup>37</sup>

<sup>36</sup> Scott Nicholson, 'A User-Centered Theoretical Framework for Meaningful Gamification', (2012) <<http://scottnicholson.com/pubs/meaningfulframework.pdf>> [accessed 5 October 2015].

<sup>37</sup> Quoted in Chris Parr, 'Why Citizen Scientists Help and How to Keep Them Hooked', *Times Higher Education*, 6 June 2013 <<https://www.timeshighereducation.co.uk/news/why-citizen-scientists-help-and-how-to-keep-them-hooked/2004321.article>> [accessed 5 October 2015].

In order to address these problems, Scott Nicholson argues that we need to develop more ‘meaningful’ forms of gamification that focus ‘on introducing elements of play instead of elements of scoring’. In other words, volunteers on transcription projects might be more engaged and motivated if the act of transcription was more playful or creative, and if it rewarded them with something of value. Crucially, this kind of meaningful gamification helps to balance the ethics of labour involved in crowdsourcing initiatives. Professional academics benefit in very tangible ways from these projects; we are, after all, paid for our time, list them on our CVs, draw upon them for publications, and use them to attract external funding. All of this is only made possible by the efforts of volunteers, who are encouraged to participate for their own enjoyment or the greater good of society. An ethically designed gamification system therefore should not be used to ‘trick’ users into doing our labour for free, but should reward them with something (whether it be an experience or a digital artefact) that they deem of genuine value.

As part of the Victorian Jokes project, we are currently building a new transcription platform to test this approach. As our users transcribe a joke and add descriptive metadata, our interface will slowly build a cartoon representation of the joke. Initially this will start with the text, but the cartoon will be updated as users provide more information about the specific joke. For example, if a user tells us that a joke features a conversation between a man and his mother-in-law, then the associated cartoon image will automatically be updated to feature a male and female character. Similarly, if a user tags the location of the same joke as a ‘Kitchen’, then the background of the image will be adjusted accordingly. Each piece of metadata added by the user will simultaneously improve both the quality of their cartoon and the quality of our database. Here is an early mock-up of what this interface might eventually look like (*Fig. 6*). Once the transcription and tagging is complete, users will have the option to share their completed comic with friends and followers on social media, or save it to a gallery on the archive’s website. As well as motivating users to contribute high-quality entries to our jokes database, the creation and circulation of the comics will publicize the project and hopefully attract new volunteers.

We are still working on developing this interface. However, in order to test the potential of releasing Victorian jokes over social media, we built the Victorian Meme Machine. This tool takes Victorian jokes from our database and automatically pairs them with illustrations from the British Library’s nineteenth-century collections. We have experimented with a range of different formats and images, each of which reimagines the jokes in different ways. Our earliest mock-ups mimicked the formatting of illustrated jokes from *Punch* magazine, which typically placed captions and lines of dialogue below a cartoon. In the case of *Punch*, the textual and

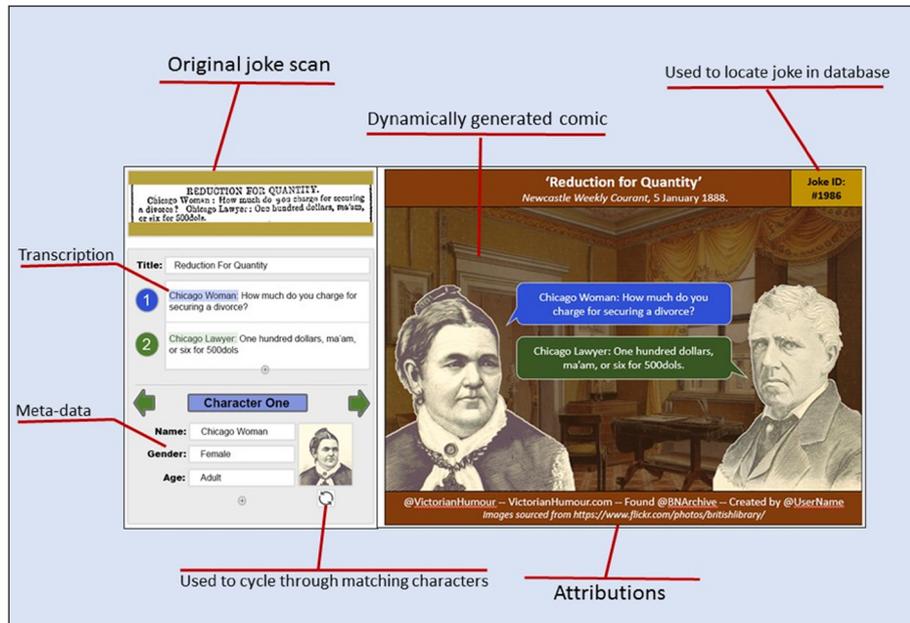


Fig. 6: Mock-up of a comic-based transcription interface.

visual elements of the joke were purpose-built to accompany one another. In our case, we needed to recycle existing images from other Victorian publications. We sourced sample images from the *Illustrated Police News* (a wonderful source of bizarre characters and situations), a million images that were extracted by the British Library Labs from nineteenth-century books and uploaded to Flickr, and the *Database of Mid-Victorian Illustration* (which has extraordinarily rich metadata for each of their images). While we initially endeavoured to find a close match for our jokes, these pictures were not designed to accompany our text and often altered its meaning in interesting new ways. Fig. 7, for example, shows the same joke accompanied by two different images. The text features a woman talking to a lawyer and enquiring about the cost of a divorce. The left-hand image features a woman sitting at a desk facing a man who could plausibly be a lawyer, which makes it a fairly literal fit for the joke. However, the image also introduces a third character, who is not present in the text: a man stands behind the woman and looks sheepishly away from the conversation. We are invited to account for his presence and factor it into our interpretation of the joke. It seems likely that this man is the woman's husband, and so his passiveness during her conversation about the cost of divorce imbues the joke with new comic dimensions. On the other hand, the right-hand version of the joke features the woman and the lawyer conversing in a very different fashion. In this case, the woman is clearly agitated; the fact that

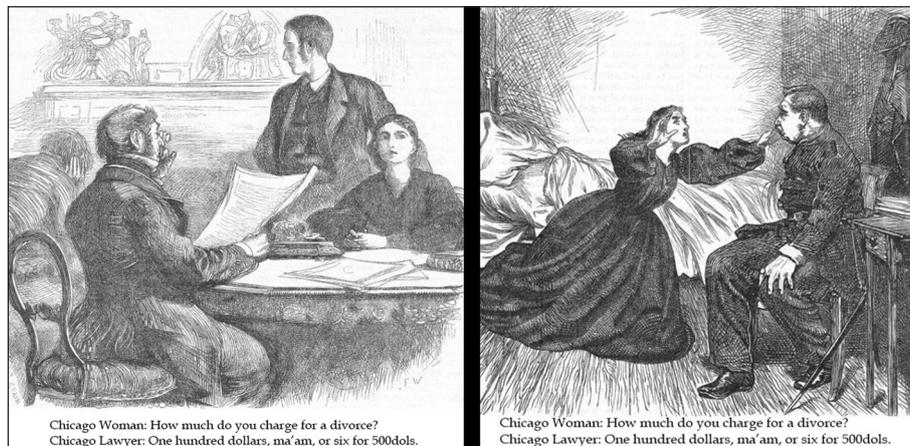


Fig. 7: *Punch*-style mock-ups of a Victorian joke paired with two different images.

she is leaping from her unmade bed indicates some kind of madness or hysteria. Why, we are led to wonder, is she so desperate to secure a divorce?

Our second batch of mock-ups adopted the style of twentieth-century comic strips and used coloured speech bubbles for dialogue. At this point, we also began to experiment with images that were even further removed from the implied context of the original joke. As the examples in *Fig. 8* show, the joke takes on another new light when the woman and the lawyer are depicted in an amorous embrace, or when the lawyer is left dangling from a rope that is being cut by her husband. The bottom example is particularly surreal and replaces the woman with a desperate-looking grizzly bear. When we tested these jokes with audiences, it was the surreal and unexpected pairings that drew the biggest laughs — the scene with the bear proved to be especially popular, and this audience feedback shaped the nature of the images that we selected for the finished version of the tool. In some respects, this process of testing and refining jokes in partnership with an audience mirrors the techniques used by Victorian editors. E. T. Raymond, who worked as a journalist in London during the 1890s, recalled how

by a process of exhaustion, the right jokes [were] reached, and by due experiment (prize competitions and the like) conducted with all the seriousness of a Home Office analysis, it is found which particular kind of joke brings the greatest happiness to the greatest number. This discovery made, the joke is made the subject of mass production, and vast stocks are poured out until the bookstall agents recommend a change.<sup>38</sup>

<sup>38</sup> E. T. Raymond, *Portraits of the Nineties* (London: Fisher Unwin, 1921), p. 305.



Fig. 8: Comic book-style mock-ups of a Victorian joke paired with four different images.

Finally, before launching into this stage of mass production, we also experimented with a design that pastiched the modern ‘memes’ that are currently very popular on social media. In visual terms, these memes usually feature white text placed over an image, with the set-up at the top and the punchline at the bottom. These images are not randomly selected, but are drawn from a relatively narrow pool of regularly reoccurring pictures — each of which is associated with a particular kind of joke. Readers who are unfamiliar with this peculiar subgenre of Internet humour may wish to consult the [Know Your Meme](#) website in order to make sense of what follows, though be warned that its contents are not entirely safe for work. *Fig. 9* shows a sample of Victorian jokes that have been mapped onto these memes in ways that are broadly faithful to their modern-day usage. Most of the images we selected for our test are conventionally used to communicate lame puns, which makes them ready-made vehicles for Victorian jokes. Some, such as the ‘[Picard Face Palm](#)’ or the ‘[Dad Joke Dog](#)’ are designed to highlight the awfulness of the joke. The ‘[Bad Joke Eel](#)’, ‘[Pun Dog](#)’, and the ‘[CSI 4 Pane](#)’ images work in a similar way, but invite us to laugh at the exuberance of the characters telling these bad jokes. The ‘[Philosoraptor](#)’ and ‘[YodaSpeak](#)’ memes are conventionally used to represent metaphysical thinking or sagacious thoughts, which makes them suited to carrying the pieces of cod philosophy that appeared in Victorian joke columns. Finally, the ‘[Anchorman](#)’ and Muppet newsreader images are usually used to carry comic news stories and headlines. These pictures are an appropriate home for Victorian jokes that masqueraded as news stories. Taken together, these formats offer an interesting alternative to the recognizably Victorian aesthetic of the first mock-ups. Without any visual or bibliographical clues, it is possible that some of these jokes might pass for genuinely modern memes. If our aim is to reintroduce Victorian jokes into the bloodstream of modern culture, then stripping them of their more obvious Victorian signifiers might give them the best chance of success. Unfortunately, the creation of the images is difficult to automate — particularly the part of the process that matches a Victorian joke to the appropriate modern meme. In many cases, it is impossible to find an image that works with the subject and structure of the Victorian text. As a result, we eventually decided to return to the Victorian-inspired designs of our earlier prototypes.

The format that we eventually adopted for the first prototype of our automated Victorian Meme Machine is displayed in *Fig. 10*. To test out our software, we selected fifty viable images from the British Library’s nineteenth-century collections. At present, these images are allocated to the textual jokes on a completely random basis. In the examples below, there is no reason why a pun about a badly conducted hotel has been paired with an image featuring an owl attacking a woman in her bed. Future versions of the tool will hopefully try to match jokes and images based on information contained within their metadata. For example, under this new system, the



Fig. 9: A selection of Victorian jokes presented in the format of modern Internet 'memes'.



Fig. 10: Images created by the 'Victorian Meme Machine'.

joke in Fig. 10 that features the word 'auctioneer' would be paired with an image that has been tagged as featuring an auction or a similar location. At present, under the random system, this joke has been paired with an image of a policeman being bitten by a female cannibal. In order to make this viable, we will need to source a large library of images and ensure that they are marked up with a detailed set of descriptive tags. The *Database of Mid-Victorian Illustration* has extremely detailed metadata for each of its images and would be a ready-made solution to this problem. Alternatively, the British Library's recent experiments with Flickr have demonstrated the power of crowdsourcing this information. Since the library uploaded a million nineteenth-century images to Flickr in December 2013, more than two hundred thousand descriptive tags have been added by members of the public.<sup>39</sup> These tags could be used by future versions of the Victorian Meme Machine to improve the pairing of texts and images. At present, our

<sup>39</sup> See the British Library's Flickr photostream <<https://www.flickr.com/photos/britishlibrary>> [accessed 4 October 2015].

software reformats the textual jokes transcribed from newspaper columns into coloured speech bubbles that appear at the top of the image. This leaves space at the bottom to provide the bibliographical details that link a joke to its original historical context. Our current design includes the title and page number of the book or newspaper from which the joke was clipped, the date that it was published, and the name of any additional authors or publications to which the joke was attributed. Future versions will include similar details for the image. These sections are larger than I would like and are in danger of detracting from the design of the main joke. Nevertheless, they are an important concession to the historical dimensions of the project and help explain the composition of the memes for audiences who are unfamiliar with it. The end results are rather hit-and-miss. By completely automating the processes of remixing, we have opened up the possibility for our software to generate surprising new juxtapositions between texts and images that would never have been created otherwise. This approach leads to more of the surreal combinations that received positive feedback from our early audiences. However, the images are often so disconnected from their texts as to make the whole thing unintelligible. As we have seen, the project's next prototype will rely on the creative decisions of our transcribers when generating the images. This should result in more intelligible jokes, but it may also sacrifice the chaotic creativity of the automated approach. The success of a joke generally rests on its capacity to surprise an audience with something unexpected; carefully curated, literal interpretations of the text may well be less effective at this than randomly generated wisecracking bears and sea monsters.

Each lunchtime, our 'Mechanical Comedian' (@VictorianHumour) posts one of these images on Twitter. The jokes are introduced with a clichéd comedy set-up line, such as 'A funny thing happened to me on the way over here...'. Most of the time, the random combination of joke, image, and set-up is rather jarring. However, sometimes the elements combine rather nicely (*Fig. 11*).

The Mechanical Comedian's performance has met with a warm, if not ecstatic, reception. At the time of writing, it has been 'telling' jokes for four months and has accumulated 256 Twitter followers. Most of the jokes receive a small number of retweets and favourites, but none have managed to go viral. When the new transcription tool is completed, this situation may improve. Users will begin to exercise their creativity and judgement over the composition of the images and will also have a greater stake in ensuring that they are shared over social media. As the project expands, we also intend to run regular 'Make This Funny' competitions. Members of the public will be invited to creatively reinterpret an expired joke in an attempt to restore its humour. These retellings might take the form of cartoons, comic strips, stand-up performances, films, animations, or any other

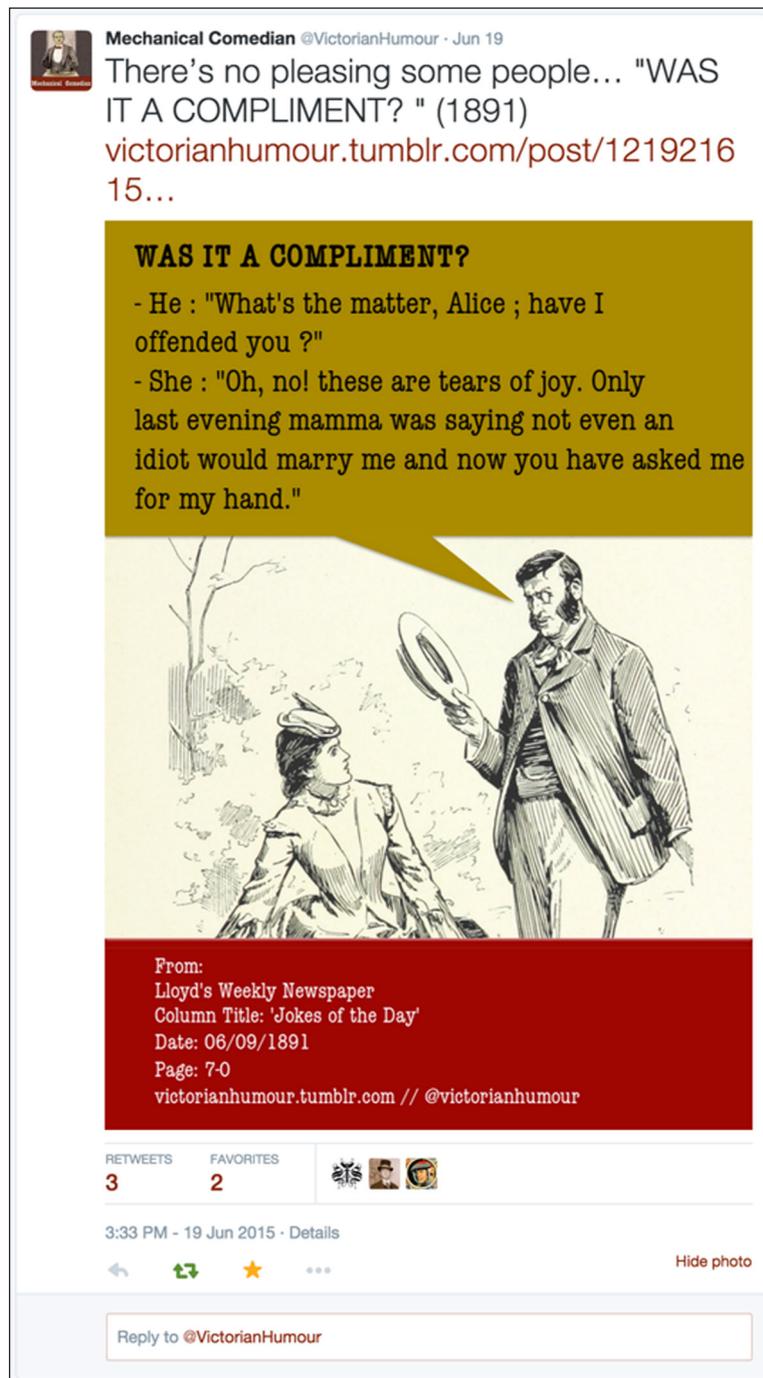
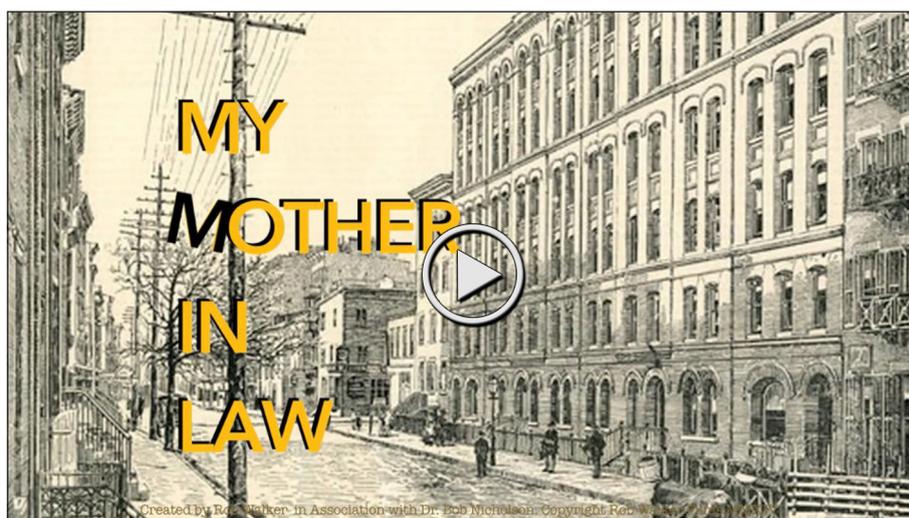


Fig. 11: A tweet by the 'Mechanical Comedian' (@VictorianHumour), 19 June 2015.

medium that users wish to experiment with. In order to test the potential of this outreach work, we sent our collection of jokes to filmmaker Rob Walker and invited him to make them funny again. He produced an animation entitled *My Mother In Law*.<sup>40</sup>



This is a parody of a 1970s sitcom that was made using original Victorian jokes and images, and presented in a style reminiscent of Monty Python. It is a curious blend that somehow works — it is certainly the closest we have come to making people laugh.

The success of Rob Walker’s animation nicely captures the ethos that underpins our wider project. Our own attempts to revitalize Victorian jokes using the Mechanical Comedian’s crude and fairly limited template have met with a relatively lukewarm response. However, when we unshackled the jokes from these restrictions and encouraged an artist to creatively engage on his own terms, the results were more exciting. The same is true of digital collections more broadly. By liberating data sets from the restrictions imposed by existing archives, we gain the chance to playfully test their possibilities; the opportunity to explore, interpret, and reuse them in interesting and useful ways. The Victorian Jokes project is a small example of how existing archives might be remixed to produce new data sets. If we are successful in our attempts to encourage crowdsourced transcription via meaningful forms of gamification, then similar approaches could be used to create a range of new archives. Imagine, for instance, how the transcription of nineteenth-century recipe columns might be incentivized

<sup>40</sup> Rob Walker, *My Mother In Law*, online animation, Rob Walker Films, 20 June 2015 <<http://www.robwalkerfilms.com/content/my-mother-law>> [accessed 5 October 2015].

by inviting users to tag ingredients. This data could then be used to automatically generate a shopping list, and then users could return with a picture and review of the dish once they have tried to make it. Sports reports, fashion columns, children's corners, stock market bulletins, and a range of other historical newspaper genres could be extracted and enriched with new metadata in equally creative ways. The Victorian jokes database represents an experimental first step along that road. There is still a lot of work to be done before its potential begins to be realized. Currently, we need more funding and resources to develop our existing tools into fully functional prototypes. In the end, by remixing the nineteenth-century archive I hope we will be able to rescue thousands of Victorian jokes from obscurity and make them accessible to academic researchers in useful new ways. But it would be a shame if their journey were to end here. As we saw with the joke about Lord Overstone, Victorian humour circulated within a vibrant culture of transnational and intertextual reprinting. By posting the jokes on social media and inviting audiences to share and reinterpret them, we open them up to an equally vibrant 'culture of retweeting'.<sup>41</sup> Perhaps, after more than a century in retirement, some of these long-forgotten jokes will once again raise a smile.

---

<sup>41</sup> Bob Nicholson, 'Tweeting the Victorians', *Victorian Periodicals Review*, 48 (2015), 254–60 (p. 257).