UNIVERSITY OF SOUTHAMPTON

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING

Electronics and Computer Science

The challenges of credibility in open news systems

by

Jonathan Scott

Thesis for the degree of Doctor of Philosophy

March 2017

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING
Electronics and Computer Science

Doctor of Philosophy

THE CHALLENGES OF CREDIBILITY IN OPEN NEWS SYSTEMS

by Jonathan Scott

The news media play a number of important roles in a modern democratic society. However, with the press under increasing pressure to deliver news more quickly and with fewer resources, some readers are turning to non-traditional sources of news. These sources are often open to contribution from ordinary citizens as well as journalists, allowing for a wider range of backgrounds and experiences to be represented. Alternative news sources feature different styles of writing, publishing schedules, and topical focus than traditional providers which can result in people holding different perceptions of these outlets.

Due to the position of the news media in society it is important to understand how these perceptions are formed and what motivates people to engage with alternative sources of news. This thesis investigates how credibility relates to levels of openness and explores methods of increasing the credibility of open news systems. I produce a landscape of citizen participation in news and find that traditional news systems do not involve citizens to the extent that the terms used to describe them imply, and that more open systems tend to lack the structure and authority usually associated with news.

An experiment is performed which shows that this 'structure and authority' is related to how credible the news source is perceived to be. A graph-based community detection algorithm is then utilised to add some level of editorial control automatically to online news discussions. This algorithm is first validated to confirm that it is able to separate news discussion contributors into meaningful groups. Then an interface is designed to present the results of this algorithm, and a study is conducted to investigate the effects of using content grouping on the credibility assessments and behaviour of news readers.

These studies find that though the interface presents a wider range of viewpoints than existing interfaces, presenting news discussions in this way does not result in a change of credibility. However there is some evidence that the changed interface may result in readers being exposed to different topics. This implies that credibility is linked more strongly to the source of the news report than the intrinsic qualities of the report.

Contents

List of Figures

List of Tables

Declaration of Authorship

I, Jonathan Scott , declare that the thesis entitled *The challenges of credibility in open news systems* and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as: Scott et al. (2015a), Scott et al. (2015b), Scott et al. (2016)

Signed:.....

Date:.....

Acknowledgements

There are many people who have contributed to the successful completion of this thesis.

First I'd like to thank my supervisor David Millard, who was very generous with his time providing advice and support. His guidance at the beginning helped me find a topic I am passionate about and his regular feedback helped ensure I stayed on track.

I am also grateful to my second supervisor Pauline Leonard who was always willing to meet with me at short notice and always offered useful guidance.

Thanks go to the many participants who volunteered their time to ensure that my research could proceed, particularly the eight interviewees who came back for more.

I would also like to thank my labmates for their stimulating discussions and for their friendship. You helped me keep this work in perspective, though without you I'd have probably finished much sooner!

Thank you to my parents who encouraged me to persue the PhD right at the start, and continued to encourage me when it got tough.

Finally, special thanks to my wife Xiyun for providing moral support throughout, for listening to me endlessly discussing this thesis, and even reading a chapter or two!

Thank you all.

In loving memory of my mother, Audrey Scott.

Chapter 1

Introduction

A preoccupation with news has existed throughout human history, presenting as verbal exchanges, written accounts, and now on television broadcasts, websites, and in mobile phone apps. In 2008 Strömbäck claimed that the news media have become "the most important source of information for most people in advanced democracies around the world". The news media fill many roles in modern democratic society: As an agenda-setter they influence the focus of public opinion (Wanta et al., 2004; McCombs, 2005), as the "fourth estate" they are expected to hold the powerful to account (Schultz, 1998), and as an information provider the media are tasked with ensuring the population are informed about the processes and decisions which concern their lives and lives of others (Anderson et al. 2013, p.9, Kovach and Rosenstiel 2007, p.10).

As the "fourth estate", the press exist outside of the three estates of parliament (the clergy, nobility, and commoners) but play an important part in the democratic process (Gentzkow et al., 2006). A free press has been described as a cornerstone of a democratic state (Lord Justice Leveson, 2012), and they are expected to critique the actions of the powerful (Schultz, 1998). However there have been a number of criticisms of the performance of the modern press in this role. One area heavily criticised is a perceived over-reliance on elite sources such as politicians and business leaders (see Davies 2011; Herman and Chomsky 1988; Bennett 1988). For example, the 2012 Leveson inquiry into the British press mentioned that "there have been far too many occasions ... when [the press'] responsibilities ... have simply been ignored" (Lord Justice Leveson, 2012).

The press have also been shown to be failing in their role as information provider. In 2013, an Ipsos MORI survey found that, among other misperceptions, British people estimated the amount of benefit fraud as 34 times higher than official estimates (Ipsos MORI, 2013). The same report also found that people thought 24% of the population were Muslim (compared to an actual figure of 5%) and 31% of the population were immigrants (compared to the official figure of 13%). A report by Ipsos MORI in 2014 found similar figures, with the public estimating that Muslims make up 21% of the

population and immigrants make up 24% (Nardelli and Arnett, 2014). With welfare and immigration regularly ranking high on election priorities (Whiteley et al., 2013) this shows that the media have not adequately informed the public of the information they need to properly participate in democracy.

Part of this issue may be explained by the lack of perspectives represented in the media. In 1972 McCombs and Shaw found that amongst local daily newspapers, national newspapers, and national news broadcasts, there existed a high degree of similarity of news agenda. This was despite this study being performed at a time when there were many independently owned news outlets. In Britain, as late as 1992, around 200 companies owned local British newspapers, whereas by 2005, just 10 corporations owned 74% of them (Davies, 2011, p.65). More recently, Davies (2011, p.52) found that 60% of the stories in four chosen "quality newspapers"[1] comprised wholly or mainly of material from newswires and public relations groups, and only 12% could be confidently attributed to a named reporter. This reveals a situation where much of the information the public receive is provided by very few sources.

In 1979, Herbert Gans called for the news media to reduce its reliance on elite sources and become more multi-perspectival, better representing other sectors of society. Over the last decade, new technologies have provided new ways for citizens to participate in the news, particularly web-based news. This has been embraced by some, for example the Guardian editor quoted by Hermida et al. (2011) as saying they want to "make lots of voices, including ones we don't agree with, heard". Others have proven reluctant to embrace these new possibilities, with Thurman (2008) quoting Independent.co.uk editor Martin King describing the users of The Independent's now closed messages boards as "a bunch of bigots who were shouting from one side of the room to the other and back again". This raises questions about the perceived value of reader participation in traditional news media.

This reluctance to relinquish control can be partially explained by journalists' "desire to preserve the status of professionals", with journalists seeing themselves as the "defining actors" in the news process (Hermida et al., 2011). This is due to concerns about the quality of content contributed by readers, with Thurman (2008) finding that journalists felt they had to "edit material to: avoid duplication, keep the standards of spelling and grammar high, select material that was newsworthy with broad appeal, and ensure balance and decency". More open systems typically do not have an editor performing these duties and this can result in output with different characteristics to that produced by the traditional news systems, including different styles of writing, publishing schedules, and topical focus. This leaves a situation where some news outlets are very open to reader participation and approach the ideal of multi-perspectival news, while others involve their audience in a superficial way while retaining overall control of the process.

[1] The Times, The Guardian, The Independent, and The Daily Telegraph

The emergence of web-based news has lead to large disruptions in the news industry. With such a large proportion of their readership moving online (according to Mitchell et al. 2016 and Ofcom 2015, 38% of Americans and 41% of UK adults use the internet for news), traditional newspapers have been quick to establish their brands online. This has taken the form of online versions of their print content as well as more experimental forms of news creation and dissemination (see Chapter 3). Also growing is the influence of alternative news systems created by groups who are not traditional news providers. This includes explicit news systems such as Wikinews and also more general-use social networks that can be used for news. In 2015, Ofcom reported that Facebook was the fifth most used news source amongst UK adults. Across these many different implementations of web-based news there exists a large amount of variation in the way that readers are permitted to contribute.

As well as having these differing levels of reader participation, online news systems also differ in how credible they are perceived to be. Some of the credibility factors are the same as the factors for traditional media: perceived levels of bias, completeness (Hellmueller and Trilling, 2012), and even whether the source agrees with the readers' pre-existing views (Metzger et al., 2003). Other factors are unique to the web, such as the form and technical affordances of the media (Chung et al., 2012). Some of the features traditionally used to judge credibility, such as author identity and reputation, are difficult to discern online, which can make it difficult to decide which sources are credible and which are not (Metzger et al., 2003). This issue has led to increasing interest in "fake news", inaccurate content using the style and structure of journalism in order to deceive.

The purpose of the work in this thesis is to investigate how openness relates to credibility. The approach taken is to first explore the forms citizen participation takes in existing online news systems, and to investigate how credible these existing systems are judged to be. The results of these studies are then used in the design and implementation of a novel interface for news discussions in order to evaluate how changes in algorithm and interface impact on readers' experiences and perceptions.

The designed interface relies on the use of community detection algorithms to group together contributions which share similar viewpoints. It is hoped that presenting the discussions using these groups will partially perform some of the tasks normally allocated to editors: avoiding duplication, stimulating debate, and ensuring balance. If this can be achieved without requiring a news editor, this could lead to systems that are both open and credible.

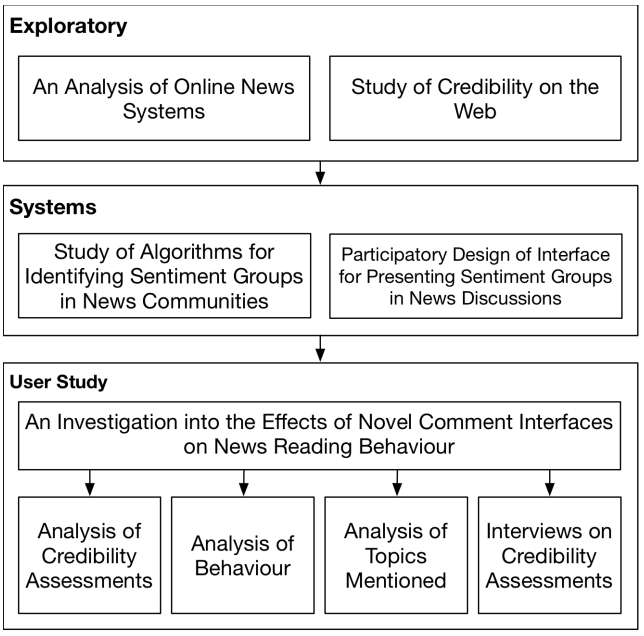


Figure 1.1: Research Framework

1.1 Research Questions

This thesis will explore news systems in terms of openness and credibility, and investigate if viewpoint grouping can be used to perform parts of the editorial role, increase the diversity of viewpoints encountered, and increase the credibility of online news systems, without reducing their openness. The questions to be addressed through this process are:

1. How might we measure openness in terms of citizen participation in the news process?
2. Does a relationship exist between the openness of a system and its credibility?
3. Will promoting a wide range of viewpoints (using viewpoint groupings) change the exposure of readers to different viewpoints, and alter the perceived credibility of a system?

1.2 Research Framework

This thesis begins with a study of existing news systems and news consumers' perceptions of them, it then proceeds with applications of algorithm and design intended to influence these perceptions, and concludes with an analysis of the effects of these changes on news consumers. The full research framework can be seen in Figure 1.1, and the individual stages are detailed below.

Exploratory

Chapter 3 investigates existing online news systems using a modified version of the framework of news production presented by Domingo et al. (2008). This framework is used to produce a spectrum of news systems ordered according to their openness to citizen participation at each stage of the framework.

The ordering of the spectrum is validated in Chapter 4 which investigates the correlation between credibility and position on the spectrum of openness. This is performed using a web-based study with 79 participants.

Systems

Having shown a correlation the research then moves on to ways in which credibility might be improved without sacrificing openness. In particular by using methods to reveal the range of viewpoints on a particular article.

Chapter 5 gives an overview of common graph-based community detection algorithms and proposes their use for diversification of news comments. The Louvain Community Detection algorithm is chosen as most suitable for grouping together contributions according to viewpoint, and this is then evaluated with a user study of 134 participants.

In Chapter 6 a participatory design process is used to create an interface to present the results of the Louvain algorithm. This is shown to result in a greater number of viewpoints presented when compared to existing interfaces.

User Study

In Chapter 6 the designed algorithm is evaluated for its impact on news consumption behaviour. This involves a study whereby participants are required to interact with several news stories, and then summarise the stories and comments and provide credibility ratings. This data is analysed to investigate if the additional structure results in different perceptions of the news, if the topics mentioned in the story and comment summaries are different due to the change in interface or algorithm, and also if there is any change in reading behaviour.

A number of these participants were invited to follow-up interviews to discuss in more depth why they made the credibility judgements that they did. These interviews are used to explain some of the results of the study.

1.3 Publications of this work to date

Parts of this work have been published as:

- Scott, J., Millard, D., and Leonard, P. (2015a). Citizen participation in news: An analysis of the landscape of online journalism. *Digital Journalism*, 3(5):737–758
- Scott, J., Millard, D., and Leonard, P. (2015b). Identifying similar opinions in news comments using a community detection algorithm. In *Social Informatics 2015*, pages 98–111. Springer
- Scott, J., Millard, D., and Leonard, P. (2016). Are open news systems credible? - an investigation into perceptions of participatory and citizen news. In *Proceedings of the 12th International Conference on Web Information Systems and Technologies*, pages 263–270

1.4 Structure of this thesis

Chapter 2 draws on the literature to define what is meant by "news", "news editor" and "citizen journalism" (as well as related terms such as "participatory journalism" and "social news"). Literature is reviewed regarding motivations behind reading and contributing to citizen journalism efforts, and news values are introduced, which will be used throughout this thesis to understand behaviours.

Chapter 3 investigates existing online news systems with a view to analysing how open they are to citizen contribution. A spectrum of openness is produced and the differences between systems on each side of the spectrum are discussed.

Chapter 4 explores this spectrum to find if the credibility of a news system is related to its openness. This concludes with some possible areas of investigation to attempt to create news systems that are more credible while retaining their openness. I explain how I will explore these possibilities by attempting to automate several of the responsibilities of a news editor.

In order to automate some of the functions of a news editor, to stimulate debate, avoid duplication, and ensure balance, there must first exist an automated way of assessing the bias in contributed content. Chapter 5 investigates a method of using voting behaviour and community detection algorithms to achieve this. This chapter covers possible algorithms, and the results of an experiment to assess the accuracy of the chosen algorithm.

Chapter 6 documents a participatory design process which produces an interface for presenting news discussions using the community detection algorithm identified in Chapter 5. This interface is compared to existing interfaces to confirm that it increases the

range of viewpoints presented. I then detail an experiment designed to investigate if the new interface, when coupled with the grouping algorithm and presented as a discussion section alongside an online news story, will change the way that people perceive the credibility of the overall story and the way they interact with it.

Chapter 7 draws together the conclusions presented in this work, evaluating the contributions made in the context of the research questions, and suggests areas for future work in the area of online citizen journalism.

Chapter 2

Background

The first part of the research framework used in this study is an investigation of online news systems in Chapter 3, but first it is important to define clearly what is meant by "news" and to outline why the news media occupies the important place in society that it does. This chapter will review existing definitions of news, and then arrive at the definition which will be used for the remainder of this thesis. The role of news in society will also be reviewed as well as common criticisms of current and historical models of news.

Next, "citizen journalism" is defined through a review of its uses in academic literature and the way that it relates to traditional journalism. The role of editors will also be investigated and news values introduced, which will be used in the rest of this thesis to aid in understanding the behaviours of editors, journalists, and news consumers.

Finally some of the major changes that are happening in the news industry due to growth of the World Wide Web are reviewed, as well as the motivations of people interacting with online news, and the risks posed by the web handing more control of the news process to the general public.

This chapter will show that though historically new technologies reduced the number of people able to make news, the emergence of web based news sources offer a response to this that has resulted in contributions from people previously excluded from the process. This offers potential to alleviate a number of common issues with the existing news media but this result is not guaranteed and techniques to manage the quality and credibility of alternative news systems must be investigated.

2.1 Evolution of News

News has existed in some form throughout human history and is seemingly ubiquitous through wildly diverse cultures and times. In Mitchell Stephens' 2007 book "A History of

News" he notes an obsession with news in places and times as diverse as fourth century Athens (where Demosthenes saw his contemporaries as pre-occupied with the news), 1712 England (where a newspaper showed concern that "shopkeepers and handicrafts spending whole days in coffee houses, to hear news and talk politicks" would lead to the breakup of families), and in 1857 Southern Africa (where a missionary wrote of men observed in a Zulu tribe "having no serious occupation, spend[ing] much of their time telling and hearing some new things"). Stephens (2007, p.14) notes that it is "difficult, if not impossible" to find a society that does not exchange news.

Stephens (2007) details the development of news through six phases: Spoken News, Written News, Printed News, Newspapers, Reporting, and Electronic News. Drawing on this framework, this section will overview each of these phases, as well as the further changes since the most recent technical innovations to disrupt the news industry: The World Wide Web and social media. This will demonstrate how the level of citizen participation has been changed by each new technology and the potential for future changes.

A complete review of the evolution of the press would necessitate the inclusion of political, social, and cultural history (Williams, 2009). Due to issues of space, only elements that are directly relevant to the study of openness and public participation in news will be considered in this brief review (for a more thorough review of the history of the press, see Stephens 2007, Williams 2009, and Curran and Seaton 2009).

Spoken News

Early news spread through word-of-mouth, by discussions in marketplaces and through travellers moving between towns. Those involved in spreading this information were not doing so with the purpose of spreading news, but instead were simply discussing the events that impacted upon their lives. This form of news allowed equal opportunities to contribute from every sector of society.

This early form of news served largely the same purpose as the modern news industry. However, due to the way that rumour and mistake can be inadvertently or maliciously introduced into spoken news, and because of the lack of control over how and which stories get passed on, political leaders soon introduced hierarchical systems whereby the powerful pass down messages to local leaders, who inform the population of their locale. These formal messengers (in the form of news criers, minstrels, bards, etc.) were primarily controlled by the powerful, and news spread by these messengers was far more likely to reach the populace, without modification or misinterpretation, than that spread informally by word-of-mouth (Stephens, 2007).

Written News

As civilizations grew and spread over ever larger areas, disseminating news by word-of-mouth became less capable of reaching the entire region, and problems of inefficiency and

unreliability became more prominent. With the development of alphabets and increasing literacy levels, societies began to use written articles distributed via messengers or postal systems. News being distributed in this way can be seen in the early Roman Empire, and in the distribution of newsletters in fifteenth century Europe (Simons, 2007).

Still, as with the news criers and bards, these methods of spreading information were available only to those of means. There was no method of mass-producing newsletters and so each had to be copied manually. Due to the expense and time required by this process, the majority of the population were not able to participate.

Printed News

Gutenberg's 15th century invention of the printing press allowed for cheap, mass production of news articles as pamphlets and newsbooks (Simons, 2007). This allowed for the dissemination of news that could reach a far greater proportion of society. These pamphlets and newsbooks would be stand-alone and typically covered only a single story, with no shared branding between issues.

Though the use of a printing press did not guarantee that the information printed was correct, it did guarantee that there were no changes made between the printing and the reading (as may be present in earlier forms of news dissemination). This allowed the reader to accept or reject the credibility of the printer without needing to consider intermediaries.

Governments seeking to influence the flow of this information placed restrictions on the ownership of printing presses, or on the information that was allowed to be printed. Outside of government control there is also evidence of self-censorship with printers choosing to avoid subjects that may upset those in power. For example, Stephens (2007) found that in 1500s England, reports of military victories were heavily reported whereas military defeats often went completely unreported.

Newspapers

Newspapers emerged in Europe early in the 17th century. These differed from the earlier newsbooks in that they were more regular, featured a greater variety of story types, and had an identity independent of the headline stories of a particular edition. There was no technical innovation that lead to the emergence of newspapers, however it did require a more developed industry to supply regular articles for publication. Whereas newsbooks would regularly publish old stories, this was no longer acceptable for newspapers, which were expected to include only recent news events.

Many of these early newspapers were associated with political parties and as late as 1870, 89% of American daily newspapers had a stated affiliation to a political party (Gentzkow et al., 2006). As with printed news, governments attempted to control who was able to print and distribute newspapers. This included the use of blasphemy and

libel laws, and also newspaper stamp duties and taxes aiming to ensure that "persons exercising the power of the press should be men of some respectability and property" (Lord Castlereagh, 1819 quoted in Curran and Seaton 2009, p.8).

Newspapers also brought with them the first emergence of "news editors" who provided a consistent voice and narrative. The impact of news editors is discussed in Section 2.4

Reporting

Early newspapers generally reported whatever news they came across. Articles were often produced based on letters received, gossip overheard, or even on the output of other newspapers (Stephens, 2007).

By the late 19th century, American newspapers were stationing special reporters in England who could communicate stories back to America via telegraph (Stephens, 2007). These reporters would collect news as their American counterparts did, but would also actively investigate stories, interview witnesses, etc. At the same time, growing cities meant that word-of-mouth was no longer capable of delivering even local news, and so newspapers, up to then primarily focused on international news, had an increased incentive to investigate local events. Over time, techniques for news investigation and reporting become more developed and became what is referred to as the journalistic method.

This period saw the rise of advertising as a method of funding news. Curran and Seaton quote some arguing that this "enabl[ed] the press to emerge as the Fourth Estate of the realm" (Curran and Seaton, 2009, p.79), though the authors themselves contend that instead of freeing the press, this instituted "a new system of press censorship more effective than anything that had gone before" (Curran and Seaton, 2009, p.81).

The 19th century also saw the emergence of news agencies: companies who sell news summaries across the world. The commercial goals of these agencies (to sell stories to as many newspapers as possible) required them to avoid partisan statements, leading to an increase in the proportion of a story dedicated to facts rather than partisan opinion (Stephens, 2007).

This professionalisation of journalism marks a large shift from the early forms of spoken news. With the emergence of reporting and the journalistic method it becomes much easier to efficiently separate journalistic news from gossip, and places more authority in the hands of those who are able to produce news in the expected format.

Electronic

The use of Radio brought another change in the news, with short, simple language becoming more popular both because of the difficulty for newsreaders to read aloud complicated language, and also to make it easier for listeners to maintain concentration and understanding (Stephens, 2007).

Providing news via the radio allowed for instant transmission of news over a large area. By the end of 1923, there were 556 stations and 400,000 receivers in use in the United States (Lewis, 1992), and by 1925, 5.5 million radio receivers were in use (Stephens, 2007). In 1933 the essayist E.B. White described radio as a "godlike presence" pervasive in American lives (quoted in Lewis 1992).

After radio came television, which for the first time allowed many people to see news events taking place in distant places. Starting with repurposed radio content with simple imagery for tiny audiences, television is now the primary source of news for people in the United Kingdom and the United States (Ofcom, 2015; Mitchell et al., 2016), and today there are a number of 24-hour news channels so that at any time of day it is possible to get news from television.

World Wide Web

The pattern appears to be that of successive technology changes enlarging the pool of people who are able to receive news, but perhaps reducing the number who can influence it, by making it more expensive or difficult to organise than what came before. For example, creating a television programme is more difficult than writing a news article, and writing a news article more difficult than spreading news by word-of-mouth.

The latest technology to revolutionize the process of producing and consuming news is the World Wide Web. This change increases the number of people who can consume the news, with articles being transmitted around the world near-instantly to anybody with access to the web, but also vastly increases the amount of people who can contribute to the news with sophisticated publishing technologies available to those same people.

In 2000, the UK Labour government said of the web that "the explosion of information has fuelled a democratic revolution of knowledge and active citizenship" (quoted in Fenton 2010, p.28). At this time the web was still a largely read-only medium where few actors created content and the rest consumed it, however technology began to fulfil this vision more completely in the mid 2000s with the advent of Web 2.0, a term popularised by Tim O'Reilly to describe web technologies which allow for more interaction than the static web pages which preceded them. Web 2.0 technologies allowing users to produce content rather than simply consume it (O'Reilly, 2005) and examples include blogs and wikis.

Alongside the growth of the web, there has been a number of innovations and experiments with news production by traditional news producers, established technology companies, and startups. These include business models such joint newspaper-website subscriptions and paywalls (see Cook and Attari 2012), specialised content sponsorship ("native advertising", see Wojdynski 2016), and entirely free ad-supported news, now challenged by the rising popularity of ad-blockers (Newman et al., 2016). For an overview of the impact of the web on UK news outlets specifically, see Saltzis (2012).

The innovations also include technical systems such as weblogs, forums, and wikis that allow for the reporting and discussion of news but greatly diminish the involvement of professional journalists. Chapter 3 will investigate the techniques being introduced by the move of news to the web, and analyse the effects of these techniques on the openness and inclusiveness of the news-making process.

Bruns (2015, p.16) argues that citizen journalism sites during this period "started out to address what they perceived as shortcomings in the mainstream media industry" but "ended up replicating many of mainstream news features". He believes this is the result of the technical structure of the systems available at the time. They required setting up a blog or content management system, and resulted in multiple individual silos of news that each needed content and promotion.

Social Media

Perhaps the most disruptive effect of the web on news is the emergence of social media platforms such as Facebook and Twitter that allow people to avoid the filtering and selection processes of news organisations and share news events in a peer-to-peer manner (though often still using them as a provider of raw information and comment). In 2015 Ofcom found that 12% of British adults use Facebook for news (the fifth most used news source among British adults), and Bruns (2015, p.7) claims that "the vast majority of local and global news organizations as well as many individual journalists now operate Twitter accounts". Some of the impact of social media on the news can be seen in its use by professional journalists for breaking news (Vis, 2013), in citizens bypassing journalists and questioning their elected representatives directly (Lee and Shin, 2012), and in journalists quoting Twitter users directly in news articles (Broersma and Graham, 2013).

The growth in the influence of citizens in the news process has lead to the questioning of the status of journalists as professionals (Singer, 2003). Social media and other forms of online news will be discussed in more depth in Section 2.8.

Examples of almost all of these types of news are still in use in modern-day societies. Though there are no longer news criers, discussion of recent events is commonplace inside workplaces and bars, and though newsbooks are no longer common, newspapers are still the favoured source of news for many people (Ofcom, 2015; Mitchell et al., 2016). The rest of this thesis will focus primarily on online news systems and on newspapers, though other forms of news will be referenced.

This section has explored the evolution of the news media with a focus on how each change has affected the range of people involved in news production. It has shown that the web offers opportunities for increasing citizen involvement in the news. In Chapter 2 I will more formally investigate the different methods used to include citizens in the news production processes of modern news systems.

2.2 Defining News

According to Stephens (2007, p.8), the word "news" has been used for at least 500 years, and "tidings" can be traced back to Old English. Kovach and Rosenstiel (2007, p.9) claim that anthropologists have found that across primitive societies, people shared "essentially the same definition of what is news". Their meaning here is that societies share the same basic news values, though, as detailed in the previous section, the way that this news is produced, distributed, and interpreted has changed over time. To focus this study requires a more structured definition of news that can categorise content as "news" or "not news".

The Concise Oxford English Dictionary defines news as "newly received or noteworthy information, especially about recent events" and The Royal Commission on the Press said "to be news an event must first be interesting to the public [...], and it must be new" (quoted in Frost 2015, p.25). These definitions depend on either an objective notion of what is noteworthy or in the public interest, or a subject for which the information is newly received.

Jarman and McClune attempted to define news in their 2007 book, beginning with a brief overview of news, they describe it as "an immensely important media form", and note a number of vague but popular descriptions of news, including "when a dog bites a man, that's not news, but if a man bites a dog, that is news", "the first rough draft of history", and "news is something someone somewhere wants to suppress".

Another definition of news presented by Jarman and McClune (2007) is that by Arthur MacEwen of the San Francisco examiner (quoted in Boorstin 1961, p.8), that news is "whatever a good editor chooses to print". This is a vague definition, but it does partially solve the problems with the Oxford English Dictionary definition of news. This definition has the "editor" deciding if the information will be newly received or noteworthy for their audience.

This view of news is consistent with that of Philo (1983) when he wrote that "news on television and in the Press is not self-defining. News is not 'found' or even 'gathered' so much as made", with Fowler (1991, p.13) when he wrote that "news is not simply that which happens, but that which can be regarded and presented as newsworthy", and with American journalist Curtis MacDougall's claim that at any moment there are billions of events that are potentially news but will not become news until "some purveyor of news gives an account of them" (MacDougall and Reid, 1968). He goes on to claim that news is "the account of the event, not something intrinsic in the event itself".

It is also consistent with Yadamsuren and Erdelez's finding in their study of 148 online news consumers that respondents' had inconsistent views on what constituted news, with some holding conceptions of news closely related to traditional media while other

respondents held a much broader view, considering "all of the Internet" to be news (Yadamsuren and Erdelez, 2011).

This definition does not account for news spread by word-of-mouth or printed prior to the development of the journalistic method in the 17th century. However, as this work focuses primarily on modern newspapers and web-based news, this definition will suffice to constrain the work in this thesis.

This document, when referring to news, will primarily use Arthur MacEwen's definition "whatever a good editor chooses to print" as this allows the classification of information as "news" or "not news". To update this to cover online news, and to reduce the subjectivity of deciding how "good" an editor is, I will modify this definition slightly to "whatever an editor chooses to publish". As this definition is dependent on an "editor", Section 2.4 will review the role of editors in news.

These definitions present news not as something which exists and is discovered, but instead as a social construction, subject to subjective decisions and institutional and cultural biases. These biases, known as news values, differ on an organisation-to-organisation and country-to-country basis, but there are many which are seen across organisations and countries. These news values will be investigated in Section 2.6.

2.3 The Role of News

As the definition of news to be used in this thesis is not specific about content or purpose, this section will review some of the roles the news media play in people's day-to-day lives. This will provide context for the chapters that follow and aid in understanding the reasons that people choose different media providers.

The media "represents the primary [...] source of information about many important events and topics" (Hall et al., 1978, p.56). As an information provider they are charged with "evaluat[ing] the policies of government and present[ing] well-informed conclusions about these key debates to the public" (Walter Lippmann summarised by Champlin and Knoedler 2006, p.138), with "provid[ing] people with the information they need to be free and self-governing" (Kovach and Rosenstiel, 2007, p.17), and with contributing to the development of democratic structures (Anderson et al., 2013, p.22). Gans (1979, p.327) believes that news is a utility of democratic society and is "inherently necessary for the proper functioning of that society".

As an agenda-setter the media influences the focus of public opinion (Wanta et al., 2004; McCombs, 2005). It has been shown that aspects of public affairs which are prominent in the news become prominent with the public (e.g. by Wanta et al. 2004; McCombs and Shaw 1972; Hall et al. 1978). The effects of this agenda-setting role on democracy are large, as it means that those who own and control large news outlets play an oversized

role in democracy. The editor of The South African daily newspaper The Sowetan once said "It simply cannot be right that, because of its dominance in the media, a minority should continue to set the public agenda" (quoted by McCombs 2013).

As the "fourth estate" the news media is expected to hold the powerful to account (Schultz, 1998), and is considered fundamental to the proper running of democracy. Breed (1955, p.328) argues that there is a "practical democratic need" for "a free and responsible press" and there are numerous examples such as the Watergate scandal (see Olson 2003) and the Edward Snowden NSA Leaks (see Greenwald 2014) that show the influence that the press can wield.

Interviews with journalists indicate that despite the decline of public trust in the media in recent decades (Jones, 2004), many see their role as one of empowering democracy. When Kovach and Rosenstiel (2007, p.19), in collaboration with the Pew Research Center, asked journalists what is the "distinguishing feature" of journalism, the most common answer given was to give people the "information they need to be sovereign". Journalistic mission statements also convey a sense of supporting democracy and public decision-making, with Kovach and Rosenstiel (2007, p.20) finding that every newspaper mission statement studied had "advancing self-government" as their primary goal.

The news media's role of gatekeepers of information - deciding what information is fit to be heard - has greatly reduced as a result of the web (Bruns 2005, p.16, Kovach and Rosenstiel 2007). As Bardoel (1996) said, journalists have moved "from an unavoidable to an avoidable link in the chain of information possession", and due to the large volume of news content produced by a huge number of outlets, journalists are no longer able to solely decide what the public should and should not know.

However, it has been claimed that though this lack of a gatekeeper results in a higher quantity of news, it also results in lower overall quality (Singer, 2003). To alleviate this, it has been argued that journalists must now fill roles of "gatewatchers" (Bruns, 2005), and of "value-adders" (Shapiro, 2010), applying judgements to the news and contextualising events for their audience (Kovach and Rosenstiel, 2007).

For many, the press also plays another role separate to these important democratic and community focused roles. For some, newspapers provide entertainment, diversion, and topics for conversation. In 1948, during a two-week strike by newspaper deliverymen, the Bureau of Applied Social Research conducted an exploratory survey of the public deprived of their newspapers, attempting to understand what makes people 'miss' a newspaper. Berelson (1948) reported on the results of this survey, finding that though "practically everyone" espoused the view that newspapers were primarily for "serious" information and public affairs, only around one third of participants could actually name a "serious" story from before the strike that they would have liked to follow up on. Instead they found that a large number of people relied on the newspaper for other uses: for radio and cinema listings, financial information, advertisements, and obituaries.

People also relied on the newspaper for relaxation, to give them conversation topics, and even for social prestige: to avoid looking "dumb and silly" when with company.

In 2001, Clyde Bentley reproduced this study by contacting newspaper subscribers who reported that their newspaper had not arrived, again attempting to understand why people 'miss' their newspaper. This study found very similar results to Berelson's (Bentley, 2001). They found that "practically everyone" expressed value in the "serious news" of newspapers but when pressed talked more about local news values, and that the same primary motivations for newspaper use that were apparent in the original study were also present in the follow up. Bentley also found that readers often had a social "ritual" whereby they would set aside specific times to read the news and discuss it with their families and friends.

2.4 The Role of the News Editor

The definition of news I am using also raises questions regarding what an editor is in this context and what their role is in the news production process. This section will review the role of an editor.

Early newspapers lacked an editor, and organised stories only by the city where the story was collected. However, by 1622 English newspaper publishers were hiring editors, with a goal of putting "newes of the same nature all together" and creating an understandable narrative from the disparate stories (Simons, 2007). Having a consistent editor over multiple editions of a newspaper also added a "voice" to the proceedings.

In 2004, Neil Thurman interviewed editors from nine British news websites regarding their views about their changing role in news, and in 2011, Hermida et al. interviewed editors from twenty four newspaper websites about the position of readers within news. During these interviews, a Washington Post editor told Hermida et al. that readers wanted "good old-fashioned journalism" and Telegraph Editor Richard Burton warned Thurman that blogs and wikis "detract from what a traditional idea of journalism is". A definition of this "good old-fashioned journalism" is not provided, but the interviews in these two studies can show what professional news editors feel their role involves.

An editor at Le Monde told Hermida et al. (2011) that encouraging debate about news events is a fundamental activity of a journalist, and a community editor at the Telegraph explained that "stimulating debate" is something they attempt to do. This is consistent with James Carey's suggestion that journalism "ought to amplify the conversation of the public" (quoted in Kovach and Rosenstiel 2007, p.18).

BBC News editors told Thurman (2008) that they aim to provide users with "a good edited read" and Hermida et al. (2011) found that some felt that it was their responsibility to "double check" user-generated news. Thurman found that journalist and editors

"felt that there was a need to edit material to: avoid duplication, keep the standards of spelling and grammar high, select material that was newsworthy with broad appeal, and ensure balance and decency".

In the context of participatory online media, Hermida et al. (2011) found that some journalists saw their role as one of "synthesizing" reports contributed by witnesses. The editor of Niuwsblad.be highlighted the importance of providing a "package" of news chosen by professionals, presenting the most important news of the day (Hermida et al., 2011), and the editor of TheSun.co.uk told Thurman (2008) that he believed that readers wanted him to "sift out content for them".

Bruns writes of professional journalists attempting to distinguish their "objective and accountable" practices from the "opinionated and partisan" citizen journalists (Bruns, 2010, p.1). This is consistent with Thurman's finding that editors felt they needed to ensure that user generated content is balanced (Thurman, 2008).

One issue with citizen journalism expressed by editors to Thurman and Hermida (2010) is that online citizen news very often involves the author personally in the story. Telegraph.co.uk Editor Richard Burton argues that "the message is the only thing that is important" and that the journalist should be "a fly on the wall" (Thurman, 2008), and former Times editor Harold Evans claims that a news story should be an unbiased account, free from the reporter's opinion (Evans, 1972).

Finally, the founding editor of the BBC News website stressed a need for an editor to mediate user contribution as they found "most people are making the same point." and claimed that typically there will "only be maybe ten points of view" (Thurman, 2008).

From the interviews by Thurman (2008) and Hermida et al. (2011), and the related literature, the primary roles of new editors appear to be:

1. Attempting to stimulate debate amongst readers (Hermida et al., 2011)
2. Editing to ensure good spelling and grammar, and ensuring decency (Thurman, 2008; Hermida et al., 2011)
3. Synthesizing and packaging content to share the most important news (Hermida et al., 2011; Thurman, 2008)
4. Ensuring objectivity and balance (Bruns 2010, p.1, Thurman 2008)
5. Avoiding duplication (Thurman, 2008)

These views of the responsibilities of editors point towards the editorial process being a particularly important part of making something "news". This leads to a tautologous definition whereby news is information produced by the process of news production.

However, this does move focus to the process of news rather than on content and this is what will be investigated in this thesis.

Under this definition much "citizen journalism" content cannot be considered to be news, as it lacks the editorial process discussed. The next section will investigate this further by looking at what is typically meant by "citizen journalism" when it is used in academic literature.

2.5 Defining Citizen Journalism

Jay Rosen defined citizen journalism as "when the people formerly known as the audience employ the press tools they have in their possession to inform one another" (Rosen, 2008), and the term "citizen journalism" has been used broadly to describe everything from bloggers producing independently researched stories to people sharing news on social networks (Robinson and Deshano, 2011). Some have attempted to make a distinction between "true" citizen journalism free from the influence of professional journalists on one hand, and citizen participation in professionally produced news stories on the other. For example Nip (2006) went as far as to identify five levels of citizen participation in journalism, in order of increasing openness: traditional journalism, public journalism, interactive journalism, participatory journalism, and citizen journalism.

A previous set of categories proposed by Lasica (2003) used the term "participatory journalism" to describe the idea of citizen participation in journalism and detailed six categories of participatory journalism: audience participation at mainstream outlets, independent news sites, full-fledged participatory news sites, collaborative media sites, other kinds of thin media, and personal broadcasting sites. Of these, personal broadcasting sites have grown so much in the years since Lasica's paper that none of the sites analysed in this thesis that would be categorised as personal broadcasting existed when Lasica proposed these categories.

Bruns (2005) took a different approach when he examined gatewatching in online news. He analysed and categorised several online news systems in terms of the gatewatching processes they use, proposing a continuum ranging from "closed news", through collaborative news sites and news-based blogs, and on to non-news sites that sometimes report news such as personal blogs and homepages.

The terms used by Nip, Lasica, Bruns, and others are used inconsistently throughout the literature and this reduces the usefulness of these terms when discussing modern journalistic endeavours. To demonstrate this, Figure 2.1 shows how these terms have been applied to a range of technical systems that support news. A total of 66 papers from 52 journals, conferences, and books were analysed (ranging from Journalism Studies to The World Wide Web Conference). A selection of uses of the terms can be seen

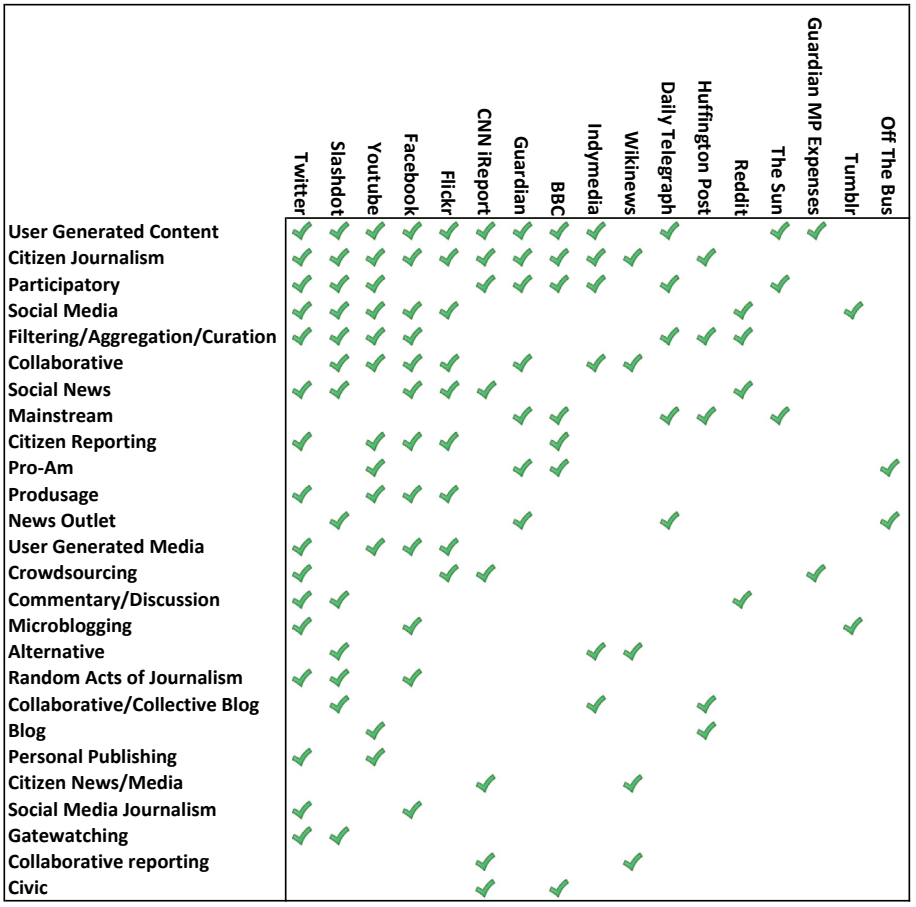


Figure 2.1: Terms used to refer to technical systems used for news, ordered by popularity on both axes

in Appendix A. The analysis is not fully comprehensive but it does show that some systems are described by a number of different terms, sometimes by terms that seem contradictory. For example, The Huffington Post was described as "citizen journalism" by Bruns and Highfield (2012), but was described as "mainstream news" by Kwak et al. (2010); and Slashdot has been described by a range of terms including "citizen journalism" (Bruns and Highfield, 2012), "participatory" (Domingo et al., 2008), "social news" (Lerman and Ghosh, 2010), "alternative" (Bruns, 2006) and "the press" (Lih, 2004). When producing the matrix I have excluded systems that were described by fewer than two terms, and terms that were used in fewer than two publications.

This confusion of terms makes it difficult to discuss and compare tools and approaches, and so Chapter 3 describes a different approach: using a model of the news production process to reveal a spectrum of participation from the more open to the more closed, comparing systems according to which users have control of each stage. I argue that this is a more sophisticated and robust view than those allowed by a set of discrete categories.

2.6 News Values

Nordenstreng (1971) claims that "news editing is specifically a process of selection", and that traditionally newsmakers have used criteria based on "news instinct" or "rules of thumb" to decide what stories would be published. These rules of thumb are known as news values, and they determine how newsworthy (or 'noteworthy', from the Oxford English Dictionary definition) a particular story is. News values enable journalists "to decide routinely and regularly which stories are 'newsworthy' and which are not" (Hall et al., 1978, p.54).

Though of course there are some differences, a number of news values are widely shared (Hall et al., 1978, p.54) and several lists have been compiled of the most common. Some of the most commonly used lists include the one by Harcup and O'Neill (2001) and, the list on which theirs is based, the one by Galtung and Ruge (1965).

Galtung and Ruge (1965) analysed 1262 press cuttings regarding crises in the Congo, Cuba, and Cyprus, and proposed twelve news values. Harcup and O'Neill (2001) analysed these values in 2001, identifying the values which were still applicable and several which had been missed by Galtung and Ruge. Harcup and O'Neill proposed their own set of ten news values. The news values from both lists are described in Table 2.1.

These news values will be used in this report to aid with understanding some of the behaviours witnessed within news communities.

2.7 Criticisms of News

Despite the important roles played by the news media, many have questioned how well journalists and editors can balance these roles against other concerns such as political and market pressures. Williams (2009) claims that "a tension between commerce and social responsibility" is at the centre of the evolution of the news media, and in 1997, one news editor said that "we are consumed with business pressure and the bottom line" (quoted in Kovach and Rosenstiel 2007, p.10)

In 1991, just 21% of Americans believed that the press cared about people, and less than half thought that the press protected democracy (Kovach and Rosenstiel, 2007, p.10), and in 2016, Pew Research Centre found that just 18% of Americans trusted the information of their national news organisations "a lot" (Mitchell et al., 2016).

This section will overview some of the major categories of press criticism, before reviewing how the move to online news will or will not respond to them. When considering these criticisms, it is important to remember the way that the news has evolved (see Section 2.1) and recognise that there has never been a period where news lacked sensationalism, bias, commercial pressures, etc. Though the criticism detailed in this section

Galtung and Ruge (1965)	**Harcup and O'Neill (2001)**	**Description**
F1 FREQUENCY		Events which fit the publication schedule of the media are more likely to be published
F2 THRESHOLD	7 MAGNITUDE	'Larger' events are more likely to be published
F3 UNAMBIGUITY		Less ambiguous events are more likely to be published
F4 MEANINGFULNESS	8 RELEVANCE	Culturally relevant events are more likely to be published
F5 CONSONANCE		Events which match the expectations of the audience are more likely to be published
F6 UNEXPECTEDNESS	4 SUPRISE	After filtering through F4 and F5, more unexpected events are more likely to be published
F7 CONTINUITY	9 FOLLOW-UP	Once something has been decided as 'news' it is likely to continue to be 'news' for some time
F8 COMPOSITION		If a particular type of news has been heavily reported, the threshold for something different to be reported is lower
F9 REFERENCE TO ELITE NATIONS	1 THE POWER ELITE	Events concerning elite nations are more likely to be published
F10 REFERENCE TO ELITE PEOPLE	2 CELEBRITY	Events concerning elite people are more likely to be published
F11 REFERENCE TO PERSONS		Events which can be understand in personal terms are more likely to be published
F12 REFERENCE TO SOMETHING NEGATIVE	5 BAD NEWS	Events with more negative consequences are more likely to be published
	6 GOOD NEWS	Very positive events are more likely to be published
	3 ENTERTAINMENT	Events including sex, showbusiness, human interest, animals, humour, entertaining photographs, or the potential for witty headlines are more likely to be published
	10 NEWSPAPER AGENDA	Events which set or fit the newspaper's agenda are more likely to be published

Table 2.1: News values proposed by Galtung and Ruge (1965) and Harcup and O'Neill (2001).

is particularly aimed at modern television and print journalism, similar criticisms can be levelled at online news (see Section 2.8) and have been directed at news throughout history (Curran and Seaton, 2009; Stephens, 2007).

The eight categories to be reviewed are the ones I see as most impacted by the move to online news, and as such most relevant to this work. However the lack of focus on other areas should not be interpreted as a judgement on their importance. Other areas of frequent press criticism include issues of privacy, context, and sensationalism.

Influence

As demonstrated in the overview of the roles of the news media, journalists have a significant amount of influence within a democracy. The Sun newspaper flaunted their influence when they famously claimed "It's The Sun Wot Won It"[1] in reference to the 1992 United Kingdom General Election. During questioning of The Sun owner Rupert Murdoch during the 2012 Leveson Inquiry, Robert Jay QC suggested that the headline implied that newspapers "were powerful and anti-democratic", though Rupert Murdoch rejected this description, claiming "we don't have that sort of power"[2]. In its conclusions, the Leveson Inquiry found that the press enjoyed too close a relationship with senior figures in the police and political parties (Lord Justice Leveson, 2012).

Anderson et al. (2013, p.8) note that due to abuse of the ideal of journalism as the fourth estate, "a great deal of care has to be exercised before concluding that a news provider is genuinely adopting a position that stands up for the interests of the people". It has been claimed that the structure of the press encourages political cynicism, presents politics as a game rather than presenting substance, and forces politicians to abide by the values of the media in order to gain visibility (Strömbäck, 2005).

Censorship

Due to this influence, those who have control over the media can gain advantages in a democracy. Attempts to control the flow of news has taken subtle forms such as taxes and stamp duties (Curran and Seaton, 2009, p.7) and licensing requirements to operate media (Stephens, 2007, p.133), as well as more overt forms such as gag orders and super-injunctions (Matthiesson, 2010) and even imprisonment (Yesil, 2014) and physical violence (Obijiofor, 2015).

The scale of these criticisms varies depending on the country where the outlets or journalists reside. Some countries have explicit protections for the media and free speech, commonly with exceptions for threatening language, indecency, information that could harm national security, etc. Countries that lack these explicit protections of the press are more often the subject of criticism. These are the countries that are ranked low on

[1] The Sun newspaper, Saturday 11 April 1992

[2] As reported in The Guardian, 25 April 2012. Accessed at `https://www.theguardian.com/media/2012/apr/25/rupert-murdoch-sun-wot-won-it-tasteless`

the World Press Freedom Index[3], which provides details on how their criteria was applied to each country. For an overview of how governments of several countries influence the press see King et al. (2013), Rahimi (2015), and Yesil (2014).

Bias

Though examples of overt censorship are relatively rare in established democracies, accusations of bias are far more common. In 2011, The Pew Research Centre found that 77% of Americans believed that "news organizations tend to favor one side" and 80% believed that that "news organizations are often influenced by powerful people and organizations" (Kohut, 2011). In 2016, when asked about political coverage, 74% of Americans thought the media favoured one side over another (Mitchell et al., 2016).

The "Propaganda Model" by Herman and Chomsky (1988) proposes five categories of filter that decide what news is able to be presented to the public, and claim that the result of these filters is a media that is biased in favour of corporate interests. According to Herman and Chomsky (1988), the categories of story that will face censorship are (in order of decreasing importance):

1. **Ownership** Stories that are damaging to the financial interests of the owners of the media.
2. **Advertising** Stories that conflict with advertiser interests.
3. **Sourcing** Stories that are overly critical of elite sources. See Section 2.7 for more specific criticism of over-reliance on elite sources within the media.
4. **Flak** Stories about topics that are targeted by pressure groups or lobbies.
5. **Fear** Stories that are not critical of public fears. The examples given by Herman and Chomsky include Communism during the Cold War and the War on Terror.

There are numerous studies which have investigated bias in the content of media outlets. For example, Groseclose and Milyo (2005) found a "strong liberal bias" in their analysis of American news outlets, and a group at the Cardiff School of Journalism found that on BBC outlets, "the ruling party has a larger share of voice" (Wahl-Jorgensen et al., 2013). There are also a number of websites dedicated to exposing the perceived bias in mainstream outlets including Biased BBC[4] and MediaLens[5]. These outlets will be discussed in Chapter 3.

Selection

[3]The latest World Press Freedom Index can be viewed at `https://rsf.org/en/ranking`
[4]See `http://biasedbbc.org`
[5]See `http://medialens.org`

Despite former Sunday Times editor Harold Evan's claim that news should be unbiased (Evans, 1972), and the stated goals of many outlets, some argue that truly unbiased news is impossible. Journalist Nick Davies claims that "the great blockbuster myth of modern journalism is objectivity" (Davies, 2011, p.111), and Herbert Gans states that even if a perfect reproduction of reality were feasible, "the mere act of reproduction would constitute a distortion of that reality". He goes on to say "objective or absolute nondistortion is impossible" (Gans, 1979, p.305).

Bias is the result not only of what is published but also of what is omitted. Lippmann (1922) said "all reporters in the world working all hours of the day could not witness all the happenings in the world" and that facts are "subject to choice and opinion". McQuail (1992) discusses the process of news production, whereby far more events happen than can reasonably be reported upon, and so the simple process of selection introduces bias.

As noted earlier in this chapter, one job of an editor is to select, of all the events that have occurred, which will produced into stories. News values (see Section 2.6) play a part in this, but there are a number of other pressures on news editors that can cause one story to be chosen over another. Reasons for selection that lie outside of the attributes of a particular story include publishers' policies (Breed, 1955, p.326) and the wish to court favour with either the general public or with societal elites (Nordenstreng, 1971).

Regardless of the goal of the journalist: to present partisan facts, a balance of multiple viewpoints, or objective truth, there must be some element of selection involved. Hall et al. (1978, p.60) comments that by exercising selectivity "the media begin to impose their own criteria" on the raw materials.

Truth

News is typically presented as factual, but there are innumerable examples of instances where the information presented by the news media was shown to be incorrect. For example, see the 1980 "Jimmy's World" story in the Washington Post which was awarded the Pulitzer prize, later returned when it was discovered that the subject of the piece did not exist (Eason, 1986), and the 1993 case of NBC Dateline admitting to "[using] incendiary devices to ensure that a fire would erupt" in order to demonstrate safety issues with particular vehicles [6]. In 2011, The Pew Research Centre found that 66% of the American public believed that "news stories often are inaccurate" (Kohut, 2011).

Many modern news outlets promote themselves as presenters of objective truth (see the BBC's commitment to impartiality [7], Wikinews' commitment to a neutral point of view [8], etc.). Though early writings sometimes used the words "truth" and "news" as if they were interchangeable (e.g. Lippmann 1920), there is much disagreement over the extent

[6] See `http://articles.latimes.com/1993-02-10/news/mn-1335_1_gm-pickup`. Accessed 2016.

[7] from BBC Editorial Values available at `http://www.bbc.co.uk/editorialguidelines/guidelines/bbc-editorial-values/editorial-values`

[8] From Wikinews policy at `https://en.wikinews.org/wiki/Wikinews:Neutral_point_of_view`

that it is possible for journalists to report the "truth" at all (Anderson et al., 2013, p.13).

Though Kovach and Rosenstiel (2007) found when interviewing journalists that more than 70% of journalists believed "there is such a thing as a true and accurate account of an event", others argue that due to the sheer complexity of human actions, presenting a single objective truth is impossible. Anderson et al. (2013, p.14) use the example of attempting to explain the rise to power of the German Nazi Party to show that explaining the actions of societies can "at best be well-evidenced possible explanations rather than 'the truth"'.

As an alternative to truth, journalists sometimes seek to provide "balance" in their reporting, providing multiple competing viewpoints. Stephens (2007, p.267) criticises this approach as "conveniently free[ing] journalists from responsibility for looking beyond competing arguments to find the truth" and states that "some events and issues [...] are unbalanced, and the effort to balance them in itself adds a kind of bias". He also points out that there are often a large number of viewpoints on an event and deciding which viewpoints are to be represented is itself a subjective decision. Kovach and Rosenstiel (2007, p.13) described the idea that "independence requires journalists to be neutral" as a myth.

Elite Sources

One aspect of the issue of selection that features prominently in news media critiques is that of an over-reliance on elite sources. Murdock (1974) notes that time and resource pressures pushing journalists to cover pre-scheduled events leads to a dependance on news sources capable of scheduling these events. Additionally, Hall et al. (1978, p.58) mentions that due to the objective of "impartiality", journalists aim to ensure that they publish statements from "accredited" sources. Becker's 1967 "hierarchy of credibility" explains that those at the top of society are seen to be credible whereas those at the bottom are seen as less credible.

Stephens (2007, p.267) claims that through the selection of these "responsible sources" the journalists "invariable demonstrate a bias", typically to those who have authority in society. Hall et al. (1978), Davies (2011), Herman and Chomsky (1988), Bennett (1988), and others believe that these issues can lead to the powerful and privileged gaining an undue amount of access to the media.

Another reason for an over-reliance on elite sources is that they are safe. Though there are numerous cases of journalists suffering due to disagreement with a government's position (e.g. the 2013 detention of David Miranda by UK authorities during the Edward Snowden NSA leaks), there are far fewer cases of individual journalists suffering for reprinting official statements from government. In the UK, this is encouraged through

the 1952 Defamation Act which protects journalists from claims of libel if they are reprinting something from a government source (Davies, 2011, p.121).

Wahl-Jorgensen et al. (2013) found that on BBC reports regarding immigration, they "included voices of immigrants and members of the public", but presented statements by politicians as "facts". Though they did also find that "ordinary people [...] are the most frequently used source type overall".

Even where non-elite viewpoints are included, stories often originate from elite sources which allows them to act as "primary definers" of the topic (Hall et al., 1978, p.59). These definers frame the problem and set the terms of subsequent debate. This occurs by setting the initial definition by which subsequent viewpoints will be compared.

Churnalism

Another issue introduced by ever-tighter deadlines, growing workloads, and shrinking budgets is the "commonly observed" issue where overworked journalists reprint the output of PR departments as news (Anderson et al., 2013, p.14), something referred to by Davies (2011) as "churnalism". As mentioned in the introduction to this thesis, in 2011 a team at Cardiff University found that newswires and PR groups contributed 60% of the stories in four analysed newspapers (Davies, 2011, p.52). Additionally, only 1% of these stories acknowledged that they had been sourced from a newswire. The researchers in this study concluded that "meaningful independent journalistic activity by the press is the exception rather than the rule".

Rock (1973) explains that institutions which generate substantial amounts of reportable activities are of particular interest to journalists. This includes courts, parliament, and sports organisations, as well as organisations who produce regular press releases. Davies (2011, p.80) reports that when political figures send a copy of a speech in advance, many journalists simply won't attend the speech and will just print what was given to them. He also bemoans the tendency of media outlets to present members of think-tanks (who produce a large amount of news content) as independent experts, despite them often being funded by special interest groups.

Multi-perspectival news

These two issues (an over-reliance on elite sources, and churnalism) taken together result in an issue of representation. In 1979, Herbert Gans performed an ethnographic study of four American national news organisations. He found that "knowns" (people known to the public, either through their own celebrity or their position in society) took between 75% and 85% of mentions in American domestic news (depending on media), while "unknowns" occupied around 20%.

Gans (1979, p.15) found that "most ordinary people never come into the news, except as statistics". In particular, he found in his study that ethnic minority voices were featured

less frequently than white voices, poor people featured less frequently than upper class people, and women less frequently than men.

Hall et al. (1978, p.55) claims that "because we [..] belong to roughly the same 'culture' it is assumed that there is, basically, only one perspective on events", and that this perspective is provided by "the culture" or the "central value system". This viewpoint, that there exists a "national consensus", neglects the difference between stratas of society, in particular the difference in power, and instead assumes that everyone has the same interests. Gans (1979, p.201) comments on this, saying that in his study, "the magazines tend [...] to universalize upper-middle-class practices as if they were shared by all Americans".

Journalism based around a national consensus accepts that disagreements will occur in society, but frames these disagreements as occurring within the bounds of the consensus. Hall et al. (1978) mentioned that this consensus is often used in news reporting, and that this both under-represents minority groups and also further re-enforces the consensus.

Anderson et al. (2013, p.15), in their investigation of quality news, state that "the more representative of the range of perspectives that are available regarding the story" the higher the quality of the article. Herbert Gans mentions that unique knowledge can be gained from different perspectives. He gives the example of poor people having knowledge about how government interacts with people at the bottom. This is knowledge that rich people cannot have by virtue of their position in society (Gans, 1979, p.311). He believes that news should "be multiperspectival, presenting and representing as many perspectives as possible"

The five key areas Gans (1979) identified where multiperspectival news would differ from the mainstream news of the time are:

1. **National** Outlets should avoid equating the government with the nation, and should report more about national corporations, unions, voluntary associations and interest groups.
2. **Bottom Up** News should include reactions not just from executives and politicians but from the people affected by events. The news agenda should also be set by a far wider range of people, including those who hold minority viewpoints.
3. **Output News** There should be more reporting of the results of policies and plans and how they have impacted different groups.
4. **Representative** There should be a greater range of viewpoints from across ideologies, ages, ethnicities, and societal roles.
5. **Service News** Outlets should provide information relevant to different groups. This means for example reporting information relevant to rich people as well as information relevant to poor people.

He predicted that the implementation of this would require news be collected from a much larger number of less easily accessible sources, the creation of new outlets to produce the increased quantity of news, and changing formats to represent many more perspectives. Ultimately, completely multiperspectival news would require a mapping of all sources and perspectives in society, which he admits may never be possible. However he does recommend smaller, more achievable changes including deliberate effort to include currently marginalised communities and regular studies of the public to determine satisfaction with the news (Gans, 1979).

In 2011, Gans revisited his proposal for multiperspectival news and evaluated the changes that have taken place in the media since 1979. He concluded that "today's news media do a better job on all five of the kinds of news I discussed" but among other suggestions, still believed that today's media needs to reduce their dependency on elite sources, provide more detail on topics other than electoral politics, and report less on unusual events (Gans, 2011).

He repeated his goals for multiperspectival journalism, claiming that the news media should "make sure that over time all the major demographic, geographic and other sectors of the country appear in the news", and that currently most news stories do not provide citizens the information they need to make political decisions (Gans, 2011).

The move to online news allows for easing a number of these concerns: distributed power can make enforcing censorship or exerting influence over the media more difficult, can increase the amount of stories (avoiding selection biases), and can increase the number of voices being presented. The next section will review the impact of online news on these issues.

2.8 Online News

In 1979, Gans lamented that "until push-button devices are invented to permit audiences to talk back instantly to top producers and editors, letter writing will have to suffice" (Gans, 1979, p.325). With the popularity of the web, this "push-button" ability to provide feedback to journalists and editors is now a reality. The web has for many people fundamentally changed their relationship with the news media. As of 2016, 38% of Americans "often" get their news online, compared to 57% who often get their news from television, and 20% from print newspapers (Mitchell et al., 2016). In the UK, 41% of adults use the internet for news, compared to 67% who use television, and 31% who use newspapers (Ofcom, 2015).

Looking at younger people, in America 50% of under-30s get news online compared to only 27% who get their news from television and 5% from newspapers (Mitchell et al., 2016), and in the United Kingdom 59% of under-25 year olds use the internet for news,

compared to 51% who use television and 21% who use newspapers (Ofcom, 2015). This suggests that the web will grow in importance as a news delivery system in the future, with newspapers facing further decline.

Many of the criticisms of the press detailed in the previous section are still valid when looking at online news sources, however the new medium brings new techniques for responding to some of these criticisms. Gans (2011) believes that the web has already increased the quantity and range of news, and has removed the space and time restrictions of traditional news media. His call for "targeted representation", for media specifically directed at particular age, interest, occupation, and religious groups has been at least partially met by the plethora of specialist news websites, and Ofcom (2015) found that online news consumers are more likely than those who get their news from TV or newspapers to rate websites highly for "offering a range of opinions", indicating that the web is having some success in increasing the range of voices presented. However Bruns (2015, p.4) points out that though citizen journalists are citizens, they "are far from representative of the overall population", and in fact are often "hardly different from those of professional journalism".

Curran and Seaton (2009, p.256) argue that as the web has become more commercial, and major corporations have created large popular websites, this has "shifted minority views to the margins", and Mitchell et al. (2016) found that while around 80% of American social network consumers will sometimes or often click on links to news from their social network, the amount of interaction declines steeply as the effort requirement increases. Just 31% sometimes or often discuss news, and 19% post their own video or photos of news events, lending credence to Gans' (2011) concern that more open news processes could just empower "the best organised and most vocal audience members" rather than the general public.

Chapter 3 will analyse existing online news systems to determine how open they are to public participation, and how the online media is performing with issues of representation, bias, and selection.

2.9 Unequal Contribution

The introduction to this thesis discussed the potential of online news systems to increase the range of voices present in the media. There is evidence that modern media has improved in its presentation of varying perspectives in the news (Gans, 2011). However, there are still many improvements to be made, in particular in media that is targeted towards a mass audience.

This is at least partially due to the demographics who access online news systems. According to Pew Research Centre's Biennial News Consumption Survey 2008 (Kohut

et al., 2008), people who consume online news regularly tend to be younger (42% of 18-29, 45% of 30-49, 27% of 50+), better educated (61% of graduates, compared to 19% of people who have no more than highschool education), and male (40% of men and 34% of women regularly use the web for news). Amongst what the Pew Research Centre called "news ranking websites" such as Reddit, the gender and age ratios are even more unbalanced, with 10% of men under 30 who use the web for news using Reddit, compared to 3% of women under 30 and 4% of men over 30. Other studies of online news consumers find similar results (e.g. Mitchell et al. 2016; Ofcom 2015).

It should also be noted that the richest segments of the world's population control the vast majority of the world's wealth. This group has the highest density of computers per head, and accounts for most of the world's online workers (Curran and Seaton, 2009, p.260). As pointed out by Curran and Seaton (2009), in 2000, USA-based content producers accounted for 83% of all pageviews. In 2015, W3Techs found that of the top 10 million websites (as listed by Alexa[9]), 53.2% used English on their homepage. The next most common language was Russian used on 6.4% of homepages (W3Techs, 2015). These findings indicate that those contributing to web-based systems are likely to be rich relative both to those in their own country and worldwide (see Norris 2001 for other impacts of this digital divide between rich and poor), and also that those who do not speak English as a first language are likely to be underrepresented.

This issue of representation is not unique to the web. In 2015, Ofcom found that among British adults, those in the AB socio-economic group[10] were more likely to consume news on television, the internet, newspapers, or radio, than those in the DE group[11].

2.10 Openness

This chapter has briefly reviewed the evolution of news over time and has highlighted the changing role of citizens in the news production process. I have established working definitions for "news" and "citizen journalism", and introduced news values and common criticisms of news which will be referred to throughout this thesis. The definition of news that will be used is "whatever an editor chooses to publish", and this chapter has identified five key responsibilities of an editor (stimulating debate, editing, synthesizing, ensuring objectivity and balance, and avoiding duplication). I have acknowledged that much "citizen journalism" content does not strictly fit this definition of news and have identified what is meant by a range of terms used to describe citizen journalism. Chapter 3 will evaluate citizen news platforms while avoiding these distracting labels.

[9] See `http://www.alexa.com`

[10] Higher/intermediate professional workers

[11] Semi-skilled and unskilled workers, casual workers, and those who depend on the welfare state for their income

Gans (1979) notes that perfect reproduction of reality is impossible, and that news can only be judged in relation to a set of standards (which themselves cannot be absolute). This thesis will compare news to the standard described by Gans (1979, p.305) as a "theory of democracy in which ordinary people are as important as public officials". Throughout this thesis, I look to online news systems to provide avenues for public contribution, and contrast systems which do with systems that have journalists in control of all stages.

For the reasons set out in this chapter, I take a normative position that news outlets should strive to represent as many viewpoints as possible, and do that by allowing the public direct access to the machinery of journalism. The next chapter will analyse existing news systems according to this ideal.

Chapter 3

An Analysis of Existing News Systems

The material in this chapter has previously been published as Scott et al. (2015a).

In 1965 Galtung and Ruge investigated the first half of what they called the "chain of news communication", the process by which a news image is produced from a world event. The complete chain had five stages: world events, media perception, media image, personal perception, and personal image. However, in the years since their work was published the news industry has changed significantly (Thurman and Hermida, 2010; Bowman and Willis, 2003). Whereas it was once the domain of journalistic gatekeepers deciding what news was fit to be heard, much news is now defined by interaction with readers. This makes the modern chain of news communication far more complex than that which existed in 1965 given the many levels of journalist and public interpretation for any single event. This new process of news production has been given many names including produsage, participatory journalism, and crowd-sourced journalism. These imply some level of public participation in news-making but are non-specific as to where this occurs and to what extent it matters to the final product.

The terms used to describe news systems (as mentioned in Chapter 2) have been applied to such a wide range of systems with differing functionality and levels of audience participation that they provide little information about those systems. This leaves one big question unanswered: How much power do citizens actually have in defining and shaping what is news? To answer this it is important to be able to determine which outlets involve the audience meaningfully and which merely claim to do so.

Ways to describe the extent of participation have been explored in other domains. For example in 1969, Arnstein proposed her "ladder of citizen participation", a typology of citizen participation looking at government instituted social programs (Arnstein, 1969).

Arnstein discussed the difference between the "empty ritual of participation" and "real power needed to affect the outcome", bemoaning the use of euphemisms such as "self-help" and "citizen involvement" as clouding the issue and making it more difficult to judge how much power is actually being given to citizens. This is a similar issue to the proliferation of terms used to describe online news systems. However, Arnstein's work does suggest a way forward, as the ladder of citizen participation focuses on what actually happens rather than the terms assigned.

Inspired by this approach, in this chapter I conduct an analysis of 32 online news systems, comparing them in terms of how much control they give users at each stage of the news production process. The objective is not to classify them, but to systematically analyse them to reveal the genuine landscape of citizen participation. In order to fully understand the process a news story goes through examples are then provided of how the online chain of news communication works by tracking several specific stories as they move through this landscape. Analysing the systems in this way will allow for meaningfully judging whether online news systems are fulfilling the ideals implied by the terms used to describe them.

This chapter will attempt to answer the first research question, "How might we measure openness in terms of citizen participation in the news process?", which will be split in to two sub-questions:

1. How much power is being given to citizens through online news systems?
2. What sort of interactions exist between those systems that are more open and those that are less?

This chapter provides an overview of the methodology, presents the 32 systems selected for study and details the eight-stage framework used for analysis. The analysis itself is then presented in the form of a matrix of citizen participation in each stage, and a spectrum constructed from the profile of each stage. Four specific stories which were reported by mainstream media but had heavy involvement of "citizen journalists" are then reviewed in order to better understand how stories move and evolve across this spectrum. Finally, this chapter concludes by evaluating results in terms of the research questions.

3.1 Methodology

The overview of terms in Section 2.5 showed that comparing systems based on the terms used to describe them does not give us meaningful comparative power. In the spirit of Arnstein's ladder, in this analysis systems are compared based on how much power they give to readers rather than on terms assigned to them or on a feature list.

In 2008, Domingo et al. performed a study of newspaper websites based on a model of the news production process. For the purposes of the work in this chapter their model will be adapted to allow for application to a wider range of modern systems. I view the distinction made by Domingo et al. between those contributing to a stage and those managing it as particularly important, taking into account Lukes' (1974) three dimensional view of power, whereby power is exercised through control of the agenda, not necessarily through individual actions or decisions.

When Bruns (2005) analysed participatory news websites, he used a simple model of news production with three stages: Input, Output, and Response. Domingo et al. (2008) built on this in their study of newspaper websites, using a six-stage model to allow for more granularity in the analysis. The stages used by Domingo et al. were Access and Observation, Selection/Filtering, Processing/Editing, Distribution, and Interpretation. These stages hase been used as the basis for the model used in this work, but some changes have been made to account for specific features of this analysis.

Firstly, some initiatives allow one community to feed into another, for example CNN iReport feeding into CNN's other properties. In these cases it is important to look at how news is distributed both within a particular outlet and how it spreads to the wider network (e.g. how CNN iReport contributions can be published on CNN's other outlets). To represent this the distribution stage has been separated into "on-site prominence" and "off-site prominence". Note that off-site prominence refers only to features of the system in question, and not on the ability of users to share news articles independently (for example on their own social media feeds). It is also important to distinguish between the initial interpretation (often the written article itself, though on CNN iReport staff can add their own interpretation which sits above the article), the ability for readers to add their own interpretation via comments, and the ability to decide the prominence of those comments. To represent this, "initial interpretation" has been made into a separate stage on the model.

Secondly, some systems allow for multiple paths through a stage, and this is represented in this analysis by splitting the cell vertically, showing the most common path on the left and the alternative path to the right. For example, prominence on Twitter is primarily decided by members (using follows and retweets), but advertisers can also pay to make their messages more prominent.

The openness of each stage was determined by reading the privacy policies and terms of use of each system, and where this was ambiguous, looking for examples of staff filtering, modifying, or censoring reader contributions. In all cases it is assumed that staff will ensure that content complies with local laws (e.g. removing copyright material) and this has not been counted as staff control unless staff are also exercising other moderation powers. The fact that at a technical level staff have overall control in all cases due to

Original	**Our analysis**	**Criteria**
Access and Observation	Access and Observation	*Who is able to submit raw data into the system?*
Selection/ Filtering	Selection/ Filtering	*Who decides which data is turned into stories?*
Processing/ Editing	Processing/ Editing	*Who produces the stories from the raw data*
Distribution	On-site prominence	*Who decides which stories users see when they are on the site?*
	Off-site prominence	*Who decides which stories get pushed to parent/external sites?*
Interpretation	Initial Interpretation	*Who provides the first interpretation seen when viewing the story?*
	Commenting	*Who is allowed to provide additional interpretation in the form of comments?*
	Comment Prominence	*Who moderates and decides the prominence of those comments?*

Table 3.1: Mapping of the stages used by Domingo et al. (2008) to our analysis stages and the criteria for analysis.

being in control of the technological infrastructure is acknowledged but is not represented on the spectrum.

Table 3.1 shows how Domingo's stages of the news production process have been expanded to provide an eight-stage framework for this analysis, and shows the criteria for openness that was applied.

3.1.1 Defining Levels of Participation

In their analysis Domingo et al. categorised each stage as closed, slightly open, moderately open, or very open, but I have chosen to be more explicit about who has control in each stage in order to avoid ambiguity in the analysis. Five participation levels are used, in decreasing order of individuals' personal investment in the system: Staff, who have an official association with the system; Paid, who pay for privileged access (e.g. advertisers, subscribers); Privileged, who without paying for it have been given extra privileges (e.g. moderators); Members who have registered for an account; and Public which includes everyone else.

3.1.2 Producing a Spectrum of News Openness

The results of this analysis were placed in a matrix mapping each system against each of the eight stages. The systems were sorted according to how open they were overall, turning the matrix into a spectrum where on the left are systems more tightly controlled by staff, and on the right are systems which offer more power to users. This was performed by sorting based on the number of stages in each system managed by staff (or unimplemented), then by the number of stages managed by paid members, then by the number of stages managed by privileged members, and so on.

The decision to treat an unimplemented stage as almost as closed as a staff-controlled stage was made in order to keep the sorting simple while giving a meaningful order to the systems. In reality whether a non-existing stage should be considered open or closed depends on which stage it is. For example, not having a filtering stage could be seen as open, but not having a commenting stage could be seen as closed. This decision could also depend on the specifics of the system being considered, and so for this work treating unimplemented stages as closed suffices to produce a meaningful spectrum.

3.1.3 Selection of News Systems

A list of news systems to be analysed was assembled by performing a literature search and an analysis of popular online websites. For the purposes of this work, "news system" means a social-technical system used for news. This means that the same technical system can appear more than once (for example Republic on Facebook and BBC News on Facebook) because they are used in a different social context.

To ensure a broad range of technical systems were covered, examples were identified of social news (Reddit, Slashdot), social media (Twitter, Facebook, Tumblr, Youtube, Flickr), mainstream citizen journalism websites (CNN iReport, Fox uReport, Al Jazeera Your Media, MSNBC FirstPerson), independent news websites (Indymedia UK, Wikinews, Wikileaks), mainstream news websites (The Guardian, BBC News, Daily Mail, The Sun, The Daily Telegraph), blogs (Wordpress, Blogger, Gawker), forums (vBulletin) and one-off newsgathering projects (The Huffington Post's OffTheBus project, The Guardian's MP's Expenses investigation).

This work is an attempt to take a representative look at citizen participation in journalism, however due to time limitations it is restricted to English language websites, primarily American and British.

3.1.4 Case Studies

When performing the analysis, it became clear that though news does typically travel sequentially through the stages, it also appears to travel across systems, for example appearing on Youtube before being discussed on Reddit and then mentioned on The Telegraph. Case studies of four stories that reached the public were used to explore this in more depth. The case studies were chosen to represent a range of different types of stories:

- "Your thoughts on Hugo Chavez" was an invitation to contribute instigated by a large media company
- "Reddit Bomb" was an investigation by an informal group who produced their own press releases
- "My tram experience" started as an individual's report of a train journey with seemingly no intent to make news
- "Transsexuals should cut it out" began with a story on the website of a traditional newspaper which was then discussed on social media.

Each of these shows a different way that news can travel through news systems. These case studies have been limited to news of short events, avoiding long-running engagements such as war. This was to make it more feasible to collect and analyse the output generated about the event, and to allow for more precisely pointing to the events that caused particular stories. Even with this reduced scope it is still not possible to collect all articles about an event and so this work focuses primarily on articles posted on the systems analysed in this chapter.

3.1.5 Building the Spectrum

The completed spectrum can be seen in Figure 3.1. Inspired by the visualisations used in Bruns' 2005 book "Gatewatching" each stage has also been plotted on the line graph in Figure 3.2 to allow for more easily spotting patterns, though whereas Bruns plotted categories of news system, I have plotted individual systems.

This section will first look at each stage in turn, and then reflect on what the analysis shows as a whole.

Access/Observation

The Access/Observation stage looks at who is able to submit data into the system. Many systems allow for either the public or members to submit information on any topic (although Fox uReport only allows data submission on pre-approved topics).

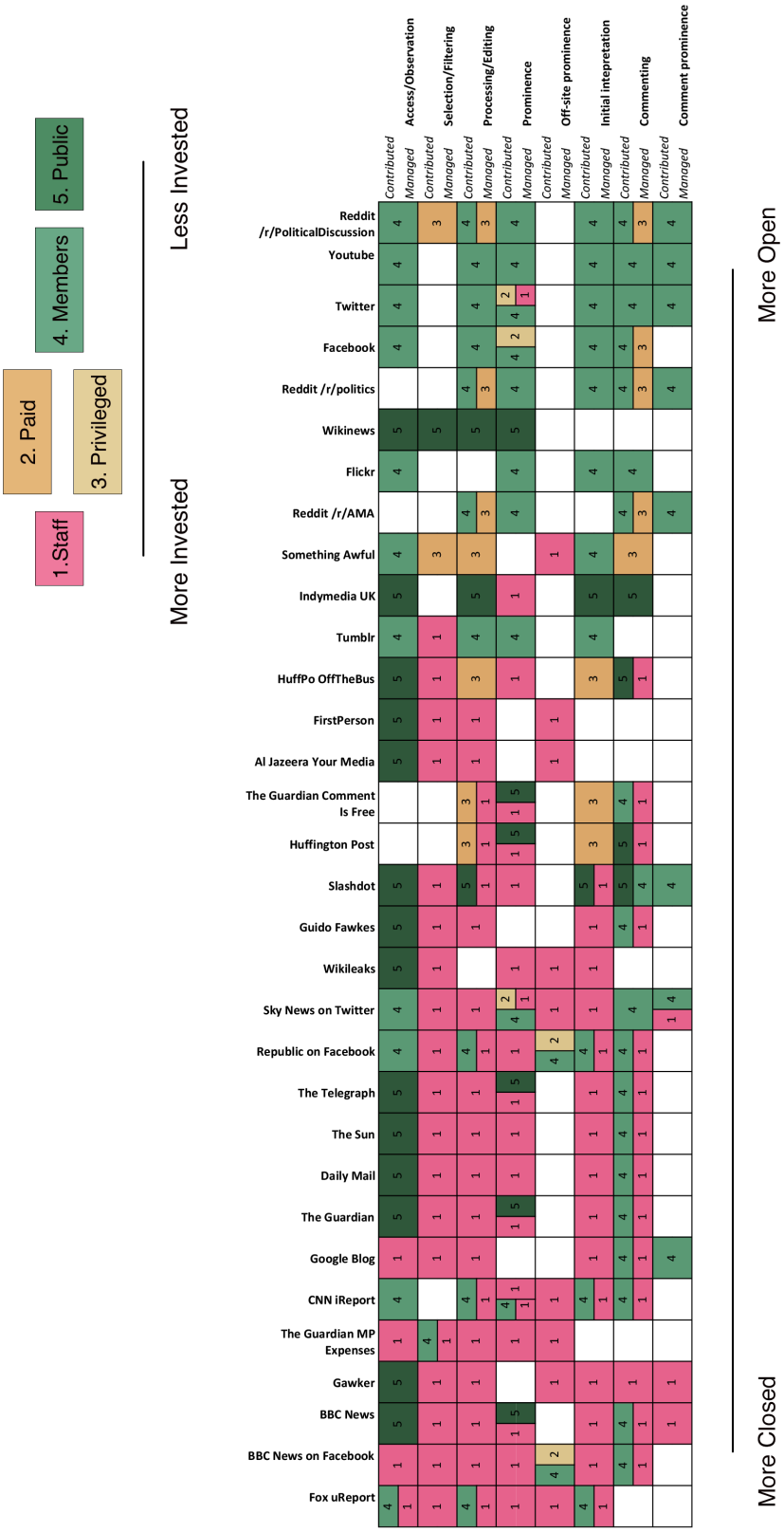


Figure 3.1: The spectrum of citizen participation in online journalism

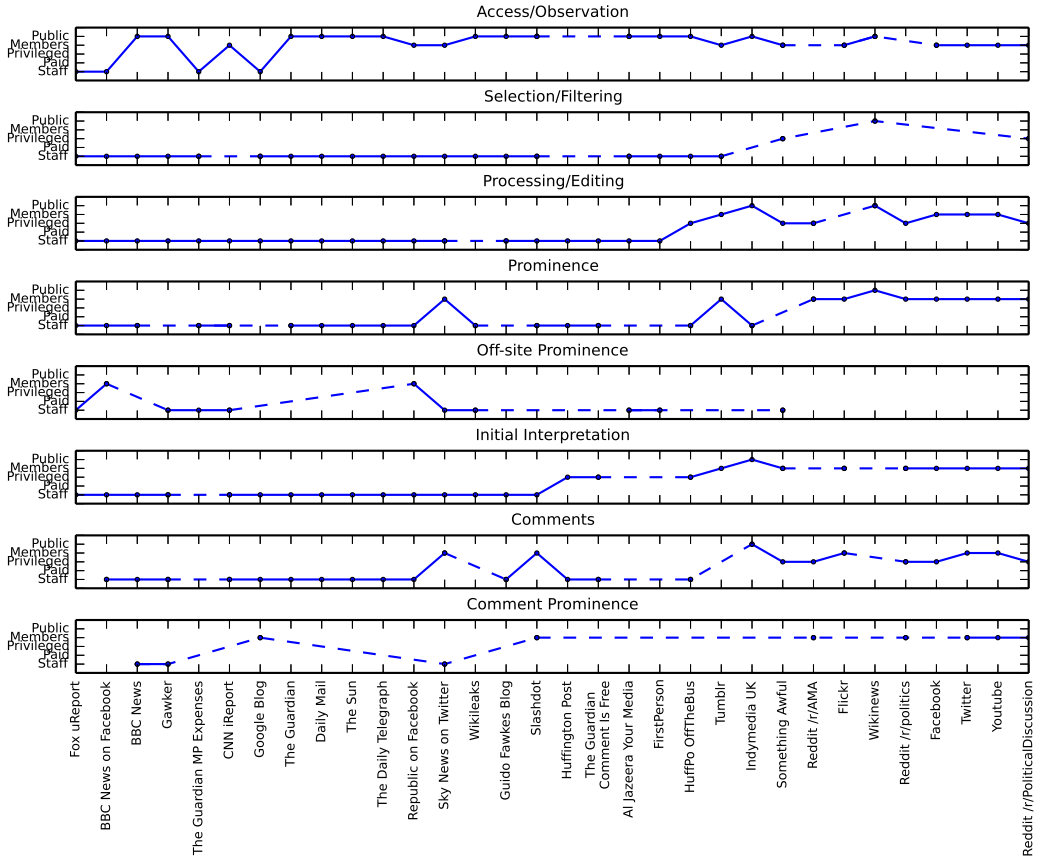


Figure 3.2: Openness by stage with systems ordered from most-closed to most-open

Selection/Filtering

At this stage it is decided which data is turned into stories. The only system considered open at the filtering stage was Wikinews, which allows the public to decide which stories get written. The PoliticalDiscussion Reddit community has privileged users (moderators) decide if a topic is suitable for discussion, and every other system which has a filtering stage has staff either in total control or managing the work contributed by others.

Processing/Editing

The processing/editing stage involves the production of stories from the data. For Reddit's politics community, where users link to existing stories rather than writing their own, this stage refers to the creation of a link rather than the production of a story.

Distribution

Distribution has been split into on-site prominence and off-site prominence.

On-site prominence represents how it is decided what users see on the site, the options seen here were: decided by members via voting or "following" sources (classified as "members"), decided by staff (classified as "staff"), decided by advertisers (classified as "paid"), or shown in chronological order (unclassified). This is different to personalisation, where prominence is altered but only for a single individual. Personalisation is not considered in this analysis.

Off-site prominence is used for sites that feed into other sites (CNN iReport, Fox uReport, etc.) where it represents how news gets selected to appear on the external site. It is also used with systems that sit atop other systems (BBC News on Facebook and Sky News on Twitter) where it represents how news spreads to the wider network. In the case of Wikileaks it refers to Wikileaks staff collaboration with mainstream news outlets.

Interpretation

The initial interpretation represents the first interpretation seen when viewing the story (the "lead" in the inverted pyramid structure of news, for an overview of the inverted pyramid see Pottker 2003). Typically this is decided by the same group that contributes to the processing/editing stage, but in the case of iReport, staff are able to optionally place their own interpretation on an article that is shown before the article itself (and so this section has been labelled as contributed by members but controlled by staff).

Staff or privileged members typically moderate comments, but in some cases (e.g. on Slashdot) this moderation power is given over completely to members. The prominence of comments is usually decided either by chronological order or member votes. BBC News is an exception in that staff members choose to feature certain comments to be shown before others (classified as "staff").

3.2 Results

Looking across the systems from most-closed to most-open shows that the access/observation stage is open even on many of the more closed systems, and the commenting stage opens up relatively early. The other stages (except for off-site prominence which will be discussed shortly) tend to stay closed for much longer. This indicates that journalists are still reluctant to give up any real gatekeeping authority, and is consistent with previous research (e.g. Hermida et al. 2011; Thurman and Hermida 2010).

Off-site prominence in most cases is either missing (because the system isn't feeding into or sitting atop another system) or is controlled by staff. The two cases where members control this stage are cases where the organisation has no choice because of the underlying system (Twitter and Facebook).

The patterns here are less clear than those shown by Bruns (2005) because of the focus on individual systems rather than broad categories. This further demonstrates the limited utility in trying to categorise these systems when there is so much variation within each category.

There is a lot of variety in the ways systems allow for citizen participation and there are very few clear groupings. The "closed news" providers are all very similar, with The Guardian, Daily Mail, Sun, and Daily Telegraph having the same openness profile, and BBC News differing only in that it has staff feature comments. The Guardian's Comment Is Free and The Huffington Post also share the same profile, whereby non-staff members contribute after being approved by staff. Al Jazeera Your Media and MSNBC First Person share a profile, both offering little more than a submission form. Social media websites Facebook, Twitter, and Youtube are very similar, and would fit well into the "Personal homepages" category in Bruns' continuum and Lasica's "Personal publishing" category. The rest of the systems however are very difficult to categorise, with too many meaningful differences to group together.

Several of the systems on the spectrum would not typically be considered to be news systems at all, but were included due to examples of them being used for news production and dissemination. In particular, sites like Youtube are not usually considered news sites but the spectrum shows that Youtube supports almost the entire news-gathering process, and suggests that perhaps there are cultural rather than technical reasons that Youtube is not usually considered to be a news site.

It is interesting that CNN iReport has a very similar profile to traditional news organisations, and Fox uReport is actually more closed than traditional news organisations. Though at first glance they appear to be citizen led systems, the analysis reveals that these systems just attempt to reproduce what is already happening on the open web within a single system that CNN/Fox can control. Kperogi made this point in his 2011 analysis of CNN iReport where he argued that "online citizen media are actually being coopted into the culture and conventions of mainstream media practices" (Kperogi, 2011).

3.3 Case Studies

It is clear that stories do not exist within a single system, but rather move and evolve as the story develops. To investigate the relationship between the more open systems and the more closed systems on the spectrum this section will now look at the spread of four specific stories in more detail. The case studies will be discussed and compared in Section 3.4.

3.3.1 Case Study 1: Your Thoughts on Hugo Chavez

When Hugo Chavez died on March 5th 2013, CNN used their iReport platform to solicit opinions on his life. They created an assignment titled Your thoughts on Hugo Chavez in English and Spanish and gave people ten days to reply. Over the next ten days, 167 reports were filed (though since then 10 have become unavailable with no explanation). Of the 157 reports still available, 95 are in English and 62 in Spanish.

Most reports posted on iReport are not vetted and CNN make no claims as to their accuracy. Once vetted by CNN, reports are stamped "CNN iReport" and are cleared to be used in other CNN outlets. Of the 157 submitted reports, 10 English and 14 Spanish reports were approved for use in CNN's other outlets. Due to the scope of this study only the English language iReports will be reviewed here.

I scraped all articles on CNN.com posted between March 5th and May 1st that were returned by a search for "Chavez", I also scraped all CNN show transcripts available on transcripts.cnn.com between these dates. With this data I searched for instances of the word "ireport", and for the titles, authors, and key parts of the description of each vetted story. The ten vetted English language iReports have been placed on Table 3.2 alongside the number of views on iReport, the number of Facebook shares (as recorded by iReport), and the places where CNN used the iReport. The usage of the English iReports can be seen on Figure 3.3, which shows the short period of time that the reports got used.

This shows that one report "10 reasons why I will not miss Chavez", an English post by Venezuelan iReporter Pancho49, received 203,000 views compared to the next highest with 4,300. Despite this, the data shows that other less popular stories were featured more prominently by CNN. Some of the stories, despite being vetted, are not mentioned in any of the materials collected. This may mean they have not been featured by CNN at all, or that they were featured on a CNN show that does not publish transcripts.

There does not seem to be any relationship between the views or recommendations an iReport receives and whether it gets vetted. One iReport, "The end of a Dictator" was vetted having received only 95 views and 2 recommendations, whereas "Who are crying for Hugo Chavez?" was not vetted despite receiving 3322 views and 377 recommendations.

3.3.2 Case Study 2: Reddit Bomb

On September 29th 2011, CNN's Anderson Cooper highlighted the "jailbait" section of Reddit (/r/jailbait), a community used for posting sexually suggestive photographs of adolescents. This led to a split in the Reddit community between those who wanted /r/jailbait shut down and those who felt that it should be permitted under the aegis of

Title	Views	Facebook Recommendations	Used
10 reasons why I will not miss Chavez	202865	38136	2 webpages
Chavez: a champion for the poor	4283	26	1 webpage, 1 TV show
Honors the memory of Hugo Chavez in Paris	3190	53	1 webpage
Rest In Peace President Hugo Chavez, You Are The Hero Of The Poor	1020	10	None
What would happen soon in Venezuela	982	20	1 TV show
Remembering the Late Venezuelan Pres. Hugo Chavez	758	24	1 TV show
Chavez Dead- Hope born	543	2	3 webpages
Most Famous Venezuelan President Hugo Chavez dies!	436	45	None
Chavez is not the way for Latin America	147	1	None
The end of a Dictator	95	2	None

Table 3.2: English iReports posted in response to Your thoughts on Hugo Chavez

free speech. The latter group included Reddit general manager Erik Martin who said to CNN "We're a free speech site and the cost of that is that there's stuff that's offensive on there"[1]. However on October 9th it was discovered that some users were using this section to trade child pornography and in response Reddit shut the community down.

On rival online community Something Awful there existed a discussion where Reddit was mocked. On February 2nd 2012 participants noticed a "preteen girls" section (/r/preteen_girls) of Reddit which had been set up to replace /r/jailbait. Much discussion then focused on this community and others like it. The participants began to discuss what they could do about these communities and one suggestion, made repeatedly, was to encourage a follow-up investigation by the media, with suggested outlets including CNN, The Huffington Post, and MSNBC.

[1] 1. As quoted on Anderson Cooper 360, CNN, September 29th, 2011

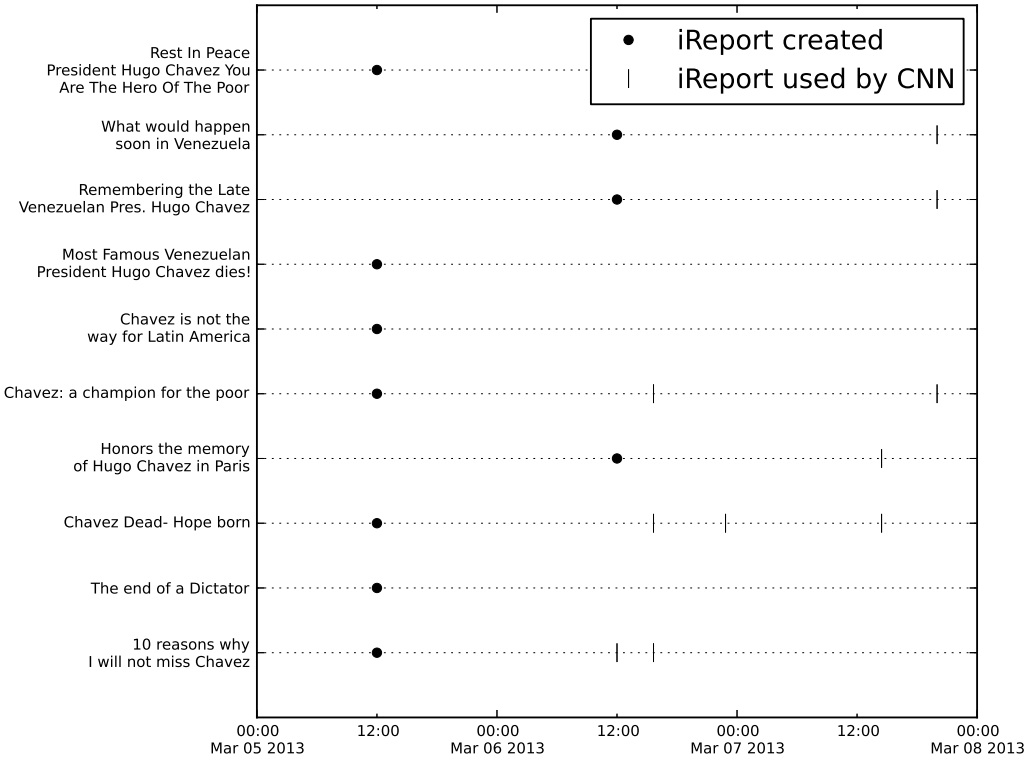


Figure 3.3: Timeline for Case Study 1, Your thoughts on Hugo Chavez

On February 12th, the first draft of a press release known as the "Reddit bomb" was created. Members of Something Awful began discussing the press release with members of the "ShitRedditSays" (SRS) section of Reddit (a community which discusses sexism, racism, homophobia, etc. on other Reddit communities). They requested that the SRS members compile screenshots of these communities and between the two communities they began sending the press release to news outlets, churches, politicians, the FBI, police, Reddit advertisers, and spreading it via social networks. Less than six hours after they began spreading the press release, Reddit introduced new rules explicitly banning sections focusing on the sexualisation of children which satisfied the majority of the contributors and discussion moved on to the response of Reddit users to the change.

News of this policy change appeared in a number of places including The New York Observer, Gawker, MSNBC, and The Huffington Post. Figure 3.4 shows the time that these articles appeared. This story was reported much more heavily in tech-focused media than in mainstream news outlets, and it appeared in only a small number of the systems on the spectrum. Many of the outlets it appeared in are online news websites running on blog platforms, and they share similar openness profiles. Not all of these systems were placed on the matrix due to issues of space, and the outlets shown in Figure 3.4 can be assumed to be similar in profile to Gawker if they are not on the matrix themselves.

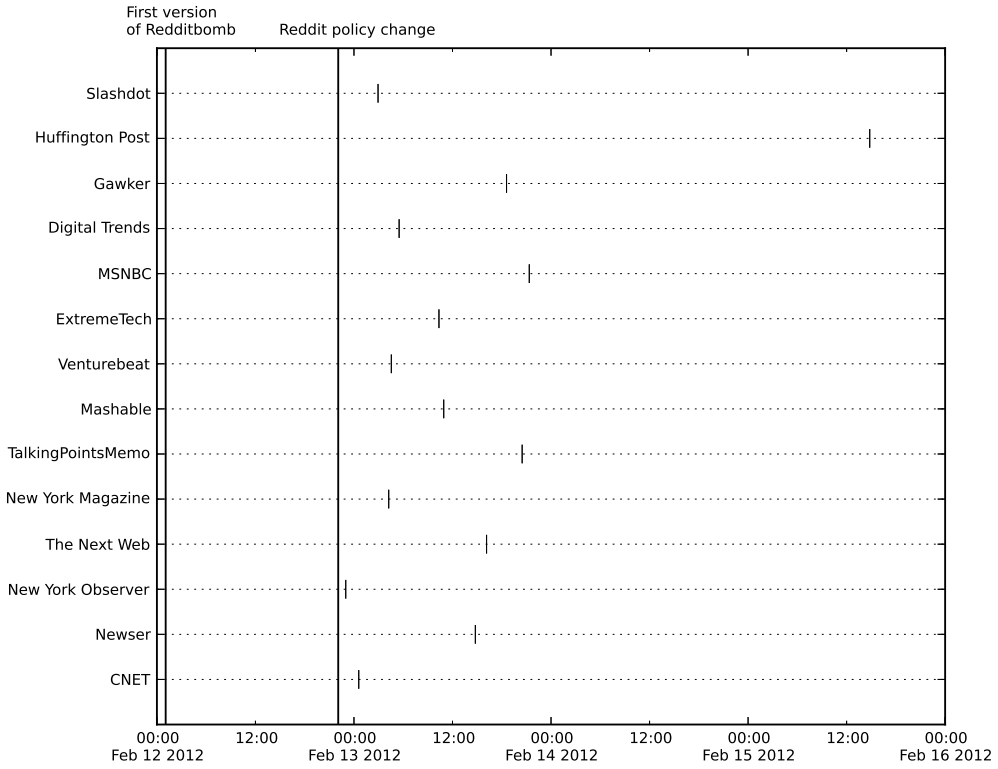


Figure 3.4: Timeline for Case Study 2, Reddit Bomb

3.3.3 Case Study 3: My Tram Experience

On November 27th, 2011, Youtube user ladyk89 uploaded a video titled "My Tram Experience" which showed a woman racially abusing passengers on a London tram. The video was discussed on Twitter where #MyTramExperience became a trending topic and discussion of the video appeared on many other websites. At 2:44am, the video was posted to Reddit under the title "I feel sorry for the kid - Racist woman on UK tram".

The Telegraph published a story about the video at 3:21pm, mentioning "many reacting angrily and calling for the woman to be arrested and charged", reflecting the sentiment in the Youtube comments without drawing attention to any comment in particular. It also quoted the description of the video provided by ladyk89 ("British woman on London transport complaining that Britain is nothing now that ethnic minorities are in her country"). The Huffington post also wrote a piece on the video at 3:44pm quoting Youtube users and linking to the video.

After the woman's arrest was announced by the British Transport Police many more outlets picked up the story, including the Guardian, the Daily Mail, and the BBC. Figure 3.5 shows when the story appeared on several of the systems on the spectrum.

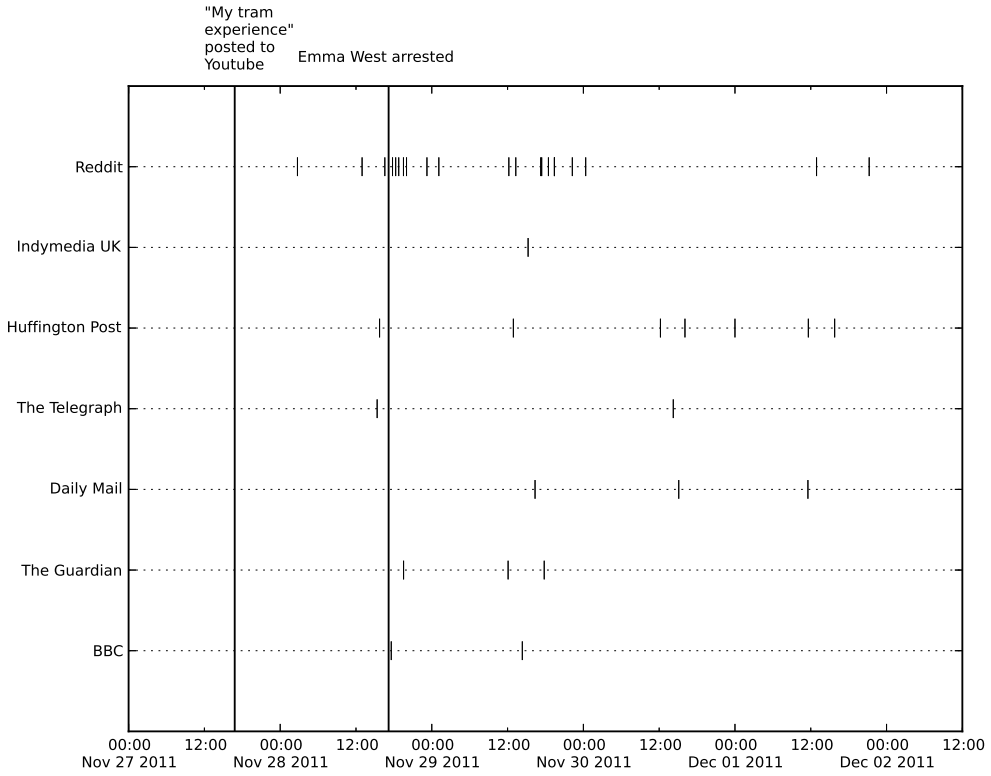


Figure 3.5: Timeline for Case Study 3, My Tram Experience

3.3.4 Case Study 4: Transsexuals Should Cut It Out

On January 8th 2013, New Statesman published an article by Suzanne Moore titled "Seeing red: the power of female anger". The article itself was about the UK recession's affects on women and towards the end of the article she wrote that women "are angry with ourselves for not being happier, not being loved properly and not having the ideal body shape - that of a Brazilian transsexual".

The New Statesman Culture Twitter account tweeted about the article at 10am and received initially positive responses. At 1:17pm, @jonanamary tweeted "Was loving this piece by Suzanne Moore on women's anger, then wham, SURPRISE TRANSPHOBIA. Ugh." followed by a tweet directed at Suzanne Moore's Twitter account explaining what she felt was transphobic. Suzanne Moore responded and the two argued for some time before others got involved.

The article was analysed on the Huffington Post and in many blogs and alternative media outlets. Suzanne Moore wrote a follow up article in the Guardian and this lead to further arguments both in the comments of the Guardian article and on Twitter, which led to Suzanne Moore deleting her Twitter account on January 11th (though she later re-joined).

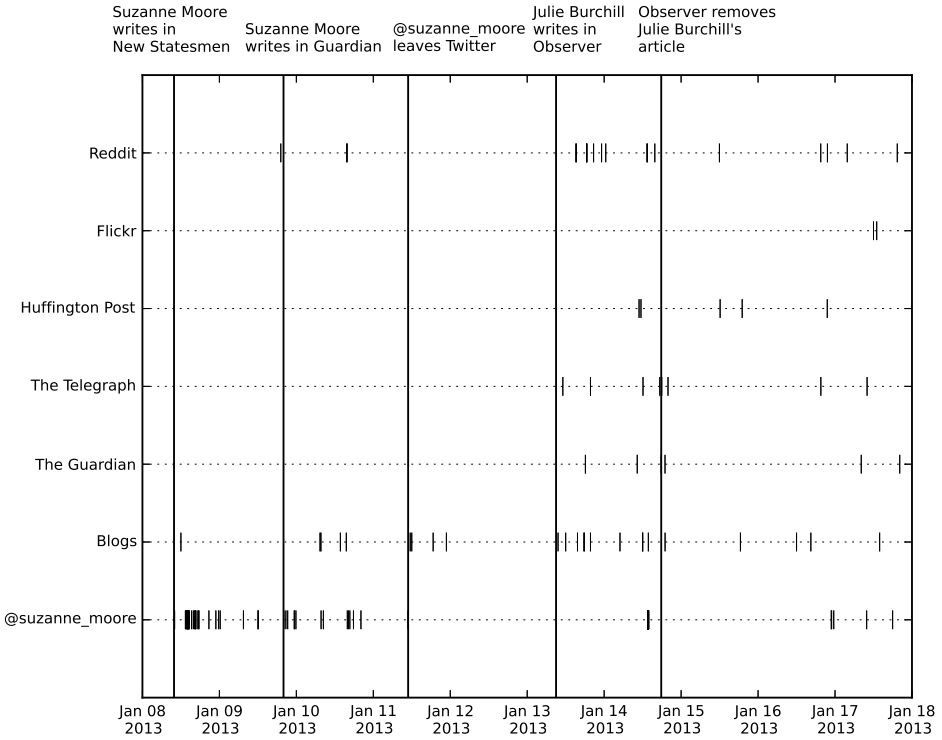


Figure 3.6: Timeline for Case Study 4, Transsexuals Should Cut It Out

On January 13th, Julie Burchill wrote an article in The Observer defending Suzanne Moore. This piece included references to transsexuals as "dicks in chick's clothing", "trannies" and "shemales", and was generally inflammatory. This caused discussion on Twitter and analysis on blogs, including blogs on The Independent and The Telegraph websites. After backlash against The Observer, the article was removed and an apology put in its place. However it was mirrored on one of The Telegraph's blogs.

A timeline of this discussion can be seen in Figure 3.6. I have included a sample of relevant blog articles, tweets, and news articles, and have excluded messages that were not referring to the articles or the discussion surrounding them. The tweets shown on the timeline are those available via Twitter's search API or via Storify.

3.4 Discussion

Whereas the spectrum allows us to compare news systems in general terms, the case studies allow us to see how they relate to one another. In the first three case studies the story moved from the more open systems to the more closed systems but in the final case study the story actually started in a relatively closed system, before being discussed on a more open system, causing further discussions on the closed system, etc. There are many examples of stories that move from open systems to closed systems, and many

organisations dedicated to analysing the output of mainstream media (e.g. MediaLens[2] and Biased BBC[3]). It is more rare for journalists to comment on the response to their stories in the way that occurred in Case Study 4.

3.4.1 News Values

This section will use the news values introduced in Chapter 2 as a means of analysing what occurred in the case studies.

In the "Reddit Bomb" case study, the forum members showed an understanding of news values when spreading the story, attempting to tailor it to each outlet. For example one user posted "I just sent the Reddit bomb to the Dallas Morning News. I also mentioned that [the administrator of the subreddit] lives in Dallas" indicating an understanding of the relevance news factor.

Due to the short amount of time that passed between the creation of the press release and Reddit's response it is unclear if the press release would have been published widely had Reddit not responded, however the fact that none of these articles mention the press release or Something Awful perhaps indicates that they see the change of policy as newsworthy in a way that the existence of the subreddits isn't. This may be because Reddit changing its rules is a simple event to report and fits more neatly into the news cycle than the on-going existence of the subreddits or the on-going campaign (Galtung & Ruge's "frequency" news value), or it may be that the policy change at Reddit evoked the "reference to elite organisations" news value (within the tech-focused news outlets that reported on this event, Reddit could be considered to be an elite organisation).

In some cases there appears to be a tipping point at which the journalistic activity itself becomes news, or triggers a newsworthy event. For example, the case of Reddit changing their policy, or the "My tram experience" story becoming much bigger once the woman was arrested. In the case of "My tram experience", the arrest gave the press an authority to quote, an effect mentioned by Davies (2011, p.120) when discussing official sources. In the case of "Transexuals should cut it out" the debate that surrounded the original story became a story in itself.

3.4.2 Routes Into The Mainstream

This analysis found no real route for citizen news to move into traditional outlets, and instead found a chaotic system where it is difficult to predict which stories will move into the public consciousness and which will not.

[2]See `http://medialens.org`
[3]See `http://biasedbbc.org`

Even looking at systems such as iReport and uReport that were established specifically to create a route for citizen news to get into the mainstream, the results seem to be very limited. In the case of the Hugo Chavez story, 24 reports were vetted out of 167 filed, not all of those vetted were used by CNN, and those which were used were used in a very small way. It is clear that this system of vetting stories and then promoting them on CNN's other, much more popular, outlets, serves to keep journalists in control of the message.

The large gap in views between "10 reasons why I will not miss Chavez" and the others, despite this apparently not being caused by CNN exposure, shows that there may be some potential for citizens to influence what news gets heard. Of course, despite being the most viewed iReport about Hugo Chavez, it still got viewed by a fraction of the 62 million unique visits CNN's news website gets each month[4]. The potential for an iReport to influence opinions needs to be judged with these comparisons in mind.

3.5 Conclusions

Due to the importance of news media in society, and the promises that are made for citizen participation in the media, it is important that we have a way of comparing different news systems that claim to have a citizen news component. However, innovation in online news progresses so quickly and in so many directions that attempts to rigidly categorise news systems can be overly constraining to the point that the topology isn't useful. I set out in this chapter to build a more sophisticated spectrum of online journalism, and in particular to compare news systems in terms of their openness. I then used case studies to investigate the relationship between open systems and closed systems. This section will summarise the results in terms of the research questions.

3.5.1 How much power is being given to citizens through online news systems?

The analysis found that there are significant differences between systems, even those that seem similar on the surface. I found very few cases where different systems have the same openness profile, though it is clear that the observation and interpretation stages are open even in many of the more closed systems and the filtering stage remains closed even in many of the more open systems.

By using the more open systems on the right side of the spectrum, citizens are able to influence the entire process of news production and distribution. Where systems do not

[4]According to CNN pressroom, January, 2013 at `http://cnnpressroom.blogs.cnn.com/2013/01/18/cnn-digital-maintains-top-spot-for-2012/`

support the entire process they can be used for the part that they do support, before moving the story to another system to continue the process.

Traditional news outlets are still tightly controlled by journalists, and offer opportunities for citizen contribution only when they can filter the contribution, or where the contribution is clearly separated from the work of journalists. The news outlets' use of social networks does not create more openness as they use these outlets only as an additional distribution channel, and even the news outlets' attempts at open news systems are still relatively closed.

3.5.2 What sort of interactions exist between those systems that are more open and those that are less?

Case studies of specific news stories were used to examine the relationship between news outlets. They showed a far more complex situation than appears from just looking at the spectrum. Specifically they showed that news stories are not constrained within the spectrum to a single system and instead move across systems as the story expands, both from open systems to closed systems and from closed systems to open systems.

This is consistent with previous findings. Davies (2011) mentioned how news organisations watch the output of other news organisations, Domingo et al. (2008) stated that the process can stop or loop and that each chain of communication can lead into other chains, and this process of the story moving from one outlet to another is very similar to what Bruns (2005) called gatewatching.

The systems towards the right (more open) side of the spectrum are systems usually labelled as "social media" rather than news outlets. Yet as discovered through the spectrum, they support all of the stages of news production. It is not clear why they are labelled differently and this may imply that they are perceived differently to the news systems on the left side of the spectrum. These systems tend to have much less structure than traditional news outlets, and looking across the spectrum from most-open to most-closed shows that as the systems become more open to public participation, the less structure they tend to have and the less they look like "news systems".

This analysis reveals a landscape of online news that is complex and intertwined, and that like the government departments when Arnstein described her ladder, mainstream news organisations don't really fulfil the promises they make of citizen participation. Citizens can look to alternative media for this kind of power, but that media lacks the structure and authority of traditional news outlets. In the next chapter I will investigate if the credibility of a news system can explain this "structure and authority". It may be that the only way to have both freedom and structure is to embrace novel techniques for news production, and this will be explore later in this thesis. Meanwhile my hope is

that this spectrum will help to shape the conversation around online news and to move it away from simple categorisation to reveal the more complex picture beneath.

Chapter 4

Study of Credibility on The Web

The material in this chapter has previously been published as Scott et al. (2016).

Previous chapters undertook a systematic analysis of news processes for a variety of different news systems, and discussed the confusion of terms used to describe them. A difference was identified between the way closed systems and open systems were labelled, but this difference was not elaborated upon. The labels range from traditional media outlets on the left of the spectrum to new social media on the right of the spectrum, and this reflects a change from systems with a strong editorial presence to those without one, though this raises the question as to what features reflect these differences.

Credibility is a measure of how "believable" a piece of information is (Fogg and Tseng, 1999) and a communicator's "competence and impartiality" (Karlsson et al., 2014). It has been considered to be one of the key elements in judging news systems (France, 1999; Cassidy, 2007; Kohring and Matthes, 2007), and the credibility of a piece of information is a key determinant of the persuasiveness of a message (Wathen and Burkell, 2002). Common factors linked to credibility include accuracy, balance, level of bias, fairness, and honesty (Hellmueller and Trilling, 2012). This chapter will investigate if the subjective difference between systems on each side of the spectrum can be explained by the perceived credibility of the systems.

This chapter will investigate the second research question: Does a relationship exist between the openness of a system and its credibility? The hypothesis is that more open systems are perceived to be less credible than less open systems.

Credibility is a subjective measure and will differ from person to person depending on their individual situation, background, and on the context of the credibility assessment. This chapter will not attempt to establish an absolute measure of the credibility of the systems but instead will investigate the relative difference in credibility between systems.

4.1 Credibility

Online news systems bring with them a unique set of challenges related to assessing credibility. With the web decreasing the cost of publishing information, the gatekeeping role of journalists is being challenged, making individuals personally responsible for assessing the credibility of much of the media they consume, and yet there is evidence to suggest that many people lack the required digital literacy skills to accurately determine the credibility of online news (Metzger et al., 2003, 2010; Flanagin and Metzger, 2000; Scholz-Crane, 1998).

4.1.1 Models of Credibility

There are a number of models of the processes used when making credibility assessments.

Petty and Cacioppo's Elaboration Likelihood Model (ELM) is a "general framework for [...] understanding the basic processes underlying [...] persuasive communications" (Petty and Cacioppo, 1986). It explains persuasion as the interpretation of central and peripheral cues.

Central cues rely on "careful and thoughtful consideration" of the content of the message whereas peripheral cues include more surface-level features such as design and usability. Peripheral cues require less cognitive effort to consider.

The Elaboration Likelihood Model suggests that when confronted with topics that are relevant to them, people are more likely to invest in the more cognitively-expensive process of examining central cues, and this results in more stable, long-term changes to attitude. However in cases where users interact with topics where they lack the requisite interest or knowledge, they are more likely to rely on peripheral cues.

In 1999, Fogg and Tseng adapted the Elaboration Likelihood Model for specific use with computer systems. They presented three approaches to judging credibility: Binary Evaluations, which involve categorising systems as either credible or non-credible, are used when users lack interest, ability, or familiarity with the topic; Threshold Evaluations categorise systems as "credible", "somewhat credible", or "non-credible" and are used when users have a moderate interest, ability, or familiarity; and Spectral Evaluations that involve far more nuanced evaluation, used when users have a high level of interest, ability, or familiarity. They suggest that threshold evaluations are the most flexible and most commonly used method of evaluating the credibility of computer systems.

Wathen and Burkell (2002) propose that when viewing online information, users will first take the less cognitively-demanding step of using peripheral cues, and only if these are decided to be credible will they make the more demanding assessment of the message. This behaviour has been shown in a number of studies (e.g. Sillence et al. 2007; Robins

and Holmes 2008; Metzger 2007), suggesting that the peripheral interpretation of surface-level credibility indicators is an important part of assessing credibility, though there is also evidence that due to the low cost of publishing online information, these surface-level judgements have very low reliability on the web (Metzger 2007).

In 2008, Hilligoss and Rieh presented their "Unifying framework of credibility assessment" which includes three levels of credibility judgement: construct, heuristics, and interaction.

The construct level refers to the user's personal definition of credibility, the heuristics level encompasses broad heuristics used to make credibility judgements, and at the interaction level the user makes use of specific central and peripheral cues.

They found that their participants conceptualised credibility in different ways: some related it to "trust" or "believability", others conceptualised it as "objectivity", and others as "reliability". Hilligoss and Rieh suggest that these differing conceptualisations of credibility result in different strategies for assessing messages. They do point out that people hold multiple definitions of credibility at once and use different definitions depending on the situation.

At the heuristic level, Hilligoss and Rieh found that participants did what was "quick" or "convenient". They found that participants used four major categories of heuristics when judging credibility: media-related, including general perceptions of the credibility of particular media (e.g. scholarly journals were generally viewed to be more credible than websites); source-related, such as known sources being seen to be more credible than unknown ones and primary sources more credible than secondary sources; endorsement-based, which involves personal recommendations, academic citations, and websites links; and aesthetics-based, which includes design and graphic elements.

The interaction level involves more specific analysis of content cues and peripheral cues, including analysis of content (comparing to the reader's existing knowledge, or comparing multiple sources of information), analysis of source peripheral cues such as affiliation and reputation, and "peripheral information object cues" which includes elements such as the language used and the presentation of information.

Fogg's Prominence-Interpretation Theory (2003) models online credibility judgements in a different way. It explains the process as as an iterative two-step process: The user first notices an element (Prominence), and then makes a judgement about it (Interpretation). These steps are repeated as new elements are noticed.

This may seem like common sense, however it does highlight the fact that an unnoticed element will have no effect on credibility. Keeping Prominence-Interpretation in mind will ensure that future experiments focus on elements that can have an impact on credibility.

Fogg (2003) identified five factors which affect prominence: the user's motivation and ability, the topic of the website, the task of the user, the experience of the user, and individual differences between users (including learning styles and literacy level). He identified the motivation of the user as perhaps the most dominant factor effecting prominence. A highly motivated user will notice more on a website than one who is casually browsing. He also identified three factors affecting interpretation: the user's assumptions, their skills and knowledge, and the context of the credibility assessment.

4.1.2 Types of Credibility

These models each offer insights into the ways that credibility decisions are taken, however they do not offer a method of assessing the perceived credibility of a particular piece of information. There are a number of methods for measuring credibility (see Section 4.1.4) and they tend to focus on one of three sources of credibility: message credibility, source credibility, and medium credibility (Metzger et al., 2003). This section will describe each of these types of credibility and how they apply to online news systems.

4.1.2.1 Message Credibility

The first area of credibility that is routinely studied is message credibility. This includes specific features of the message content, including information quality; language intensity; and message discrepancy (Hellmueller and Trilling, 2012).

As this work is interested in the relative quality of individual news websites, rather than different messages within a single website, message credibility will not be relevant to this study.

However, it has been shown that the credibility of one piece of content influences the perceived credibility of other content on the same page (Thorson et al., 2010). Though the more closed systems investigated do have content from multiple authors on a single page, this is far more common on the more open systems. This may impact on the credibility of the systems to the right of the spectrum and should be taken into account.

4.1.2.2 Source Credibility

Source credibility examines the personal characteristics of the message source that influence credibility judgements. Examples of factors that contribute to source credibility include expertise, trustworthiness, and sociability (Wathen and Burkell, 2002; Self, 1996).

Tseng and Fogg (1999) identified four types of source credibility: Presumed Credibility includes factors based on the assumptions of the receiver; Reputed Credibility is based on labels, such as "Journalist", "Activist", or "Doctor"; Surface Credibility is based on superficial characteristics such as appearance; and Experienced Credibility is the most reliable of Tseng and Fogg's four types of credibility, being formed based on experience with the source over time.

On the web, the influence of source credibility is difficult to measure as the source of a message is often misattributed: to website owners, sponsors, and even to web designers (Metzger et al., 2003). Many online news sites have content from multiple sources and the impact of each individual source on overall credibility is difficult to extract. However, when correctly identified, the reputation of the source is a key element of the credibility of a piece of information.

Schweiger (2000) found that news organisations that have credibility in traditional media may have their web presence judged more credible than those that lack this traditional media presence. This can be explained as a feature of Tseng and Fogg's "experienced credibility" (Tseng and Fogg, 1999) and an assumption that the organisation maintains similar editorial standards on the web as they do in non-web media.

Sundar and Nass (2001) found that the credibility of what they called the "selecting source" was also important when judging source credibility. The selecting source is the entity that chose to share the content rather than the one that initially collected or created it. Sundar and Nass (2001) found that users viewed information selected by other users to be more credible than information selected by an expert, though Sundar et al. (2007) found that the credibility of the original source is still important. Schmierbach and Oeldorf-Hirsch (2012) found that this effect of the selecting source increasing the credibility of a piece of content did not apply to content shared on Twitter, and posited that this may be due to the Twitter platform's reputation as a place for "celebrities and shallow posts".

Some have identified design features of websites and message quality as potentially important to users making judgements of source credibility (e.g. Metzger et al. 2003; Fogg and Tseng 1999; Olaisen 1990; Slater and Rouner 1996), what were refered to by Olaisen (1990) as "technical" qualities (as opposed to the "cognitive" factors). This would be part of Tseng and Fogg's "surface credibility".

4.1.2.3 Media Credibility

Media credibility concerns the credibility of a communications medium. Traditionally, much media credibility research has concerned the relative credibility of television news and newspapers, with newspapers typically being perceived to be less accurate and more biased than television (Metzger et al., 2003). Reasons for this include that television

can report live breaking news which proscribes authority and importance (Chang and Lemert, 1968), that television allows consumers to see what is happening rather than it being described (Carter and Greenberg, 1965; Gaziano and McGrath, 1986), and because TV news typically covers fewer stories than newspapers they are less likely to make mistakes which reduce confidence (Wilson and Howard, 1978), or to push dissonant messages which increase the appearance of bias (Carter and Greenberg, 1965).

Some of these differences between television news and newspapers are also present when comparing newspapers and online news. Online news is able to react to ongoing events even quicker than television news, and with the prevalence of camera-enabled mobile phones and modern web technologies, online news systems are able to show videos of events promptly, sometimes even as they are occurring. These features may lead to websites being perceived as more credible than newspapers, but other factors will lead to the web having less credibility. These include the fact that on the web there is much more information, on many more topics, from many more people and points of view, than there is in newspapers. This increases the chances of negative message and source credibility factors influencing judgements, as readers are likely to be presented with non-credible sources and incorrect information. This can lead to the medium appearing biased.

This means that depending on the weighting attributed to each factor, the web as a whole may be seen as less credible, equally credible, or more credible than newspapers. Research is available to support each of these positions. For example see Johnson et al. (2007) and Johnson and Kaye (1998) which found online information to be more credible than offline information, Kiousis (2001) and Flanagin and Metzger (2000) which found online information to be equally or more credible than television but less credible than newspapers, and Schweiger (2000) which found online information to be less credible than both newspapers and television.

Which type of credibility ought to be focused on is unclear due to the inconsistent use of terms in the literature. The systems placed on the spectrum could be classed as either sources or as media, and the decision will affect which measures are used to analyse the systems. As mentioned by Hellmueller and Trilling (2012), The Yale Group Communication Research Program defines “sources” as individual persons, institutions or magazines, whereas McCroskey and Young (1979) defined this as individual communicators. The systems on the spectrum could be classed as source institutions under the Yale Group definition but would not fit McCroskey and Young’s definition. The systems could also be classed as news media, with the individual contributors acting as sources.

This is a particularly important distinction due to the types of contribution made to the systems being analysed. Though some systems present content in a similar manner to traditional newspapers (particularly those systems towards the “closed” side of the spectrum), others (most obviously Facebook and Twitter, though also sites such as

Comment Is Free by The Guardian) allow contributions from such a wide range of people that treating it as a single "source" would not accurately reflect the diversity of contribution. For this reason, this study will consider the technical systems to be news media, and will use media credibility measures to investigate the differences on each side of the spectrum.

4.1.3 Spectrum of Credibility

Having built a spectrum of openness in Chapter 3. This section will overview available literature on the credibility of the analysed systems to build an idea of how openness relates to credibility.

From the more-closed systems typically labelled as traditional, Flanagin and Metzger (2007) found that traditional online news sources ranked higher than other news sources in credibility, and Melican and Dixon (2008), when investigating credibility and racism in online news, found that non-traditional internet news sources were perceived as less credible than all other sources of news. However, Kim and Johnson (2009) found that politically-interested South Koreans viewed independent online news sites as more credible than traditional media websites.

In the centre of the spectrum, blogs have been shown in some cases to be rated as trustworthy, though primarily by people already skeptical of mainstream media or regular users of blogs, with non-users holding lower opinions of blogs' credibility (Johnson and Kaye, 2004, 2009; Johnson et al., 2007; Choi et al., 2006), but Schmierbach and Oeldorf-Hirsch (2012) found that overall, blogs were less credible than mainstream news.

For the more open systems, Schmierbach and Oeldorf-Hirsch (2012) compared information presented via a news organisation's Twitter feed to information presented directly by the news organisation, and found that "Twitter is seen as a less credible source that may or may not present a less credible message". Marchionni (2013) explored the conversational features and credibility of news on Twitter, Wikinews, and what Thorson and Duffy (2006) called "collaborative news", where professional journalists utilise many quotations and explicitly report from the perspective of the audience. They found that their participants rated collaborative news as more credible than Twitter, which in turn was more credible than Wikinews.

The Pew Research Centre's Biennial News Consumption Survey 2008 found that Americans find "most traditional news sources - print, TV and radio - at least somewhat credible". They also found that online news sources were "viewed with more skepticism", and "none is viewed as highly credible by even a quarter of online users able to rate them" (Kohut et al., 2008).

This provides some evidence that more closed systems tend to be viewed as credible, and more open systems tend to lack credibility. However, there is conflicting data, and due to the inconsistencies in scales used, the differing definitions of 'credibility' used (see Hellmueller and Trilling 2012), and the lack of data available for many of the systems on the spectrum, overall conclusions regarding the relationship between openness and credibility cannot be drawn from these previously published studies.

Instead, in this chapter I will perform my own study of the perceived credibility of these news systems to ensure that a consistent criteria is applied and the results can be directly compared to the spectrum of openness presented in Chapter 3.

4.1.4 Measuring Credibility

There have been a number of proposed ways of measuring the perceived credibility of a piece of media. Hellmueller and Trilling performed a meta-analysis of credibility research between 1951 and 2011, finding inconsistent use of concepts and a "proliferation of questionable items and scales" dominating credibility research (Hellmueller and Trilling, 2012). They found that of the 85 scales they investigated, only 16 replicated previously used scales. This makes deciding on the scale to use in this study difficult.

One scale which was used often, though can not be considered a standard, is that by Meyer (1988). Meyer validated the twelve factors previously proposed by Gaziano and McGrath (1986) and produced five factors which represent credibility. These were shown by West (1994) to validly and reliably measure credibility.

In Hellmueller and Trilling's meta-analysis, Meyer's credibility index was the most commonly used set of factors, being used 6 times to measure media credibility (out of 29 in total), 3 times for source credibility (out of 21), 2 times for message credibility (out of 8), and 3 times in other studies (out of 10).

In the absence of an accepted standard, this study will use the Meyer (1988) credibility scales, the factors of which are:

1. Unbiased/Biased
2. Fair/Unfair
3. Tells the whole story/Does not tell the whole story
4. Accurate/Inaccurate
5. Can be trusted/Cannot be trusted

4.2 Methodology

In order to measure the relationship between position on the spectrum proposed in Chapter 3 and the perceived credibility of the news systems, a web-based study was employed where participants were asked to rate the systems according to Meyer's five factors (ERGO #12819).

To ensure that the experiment is understandable, two participants were asked to individually complete the study using a talk-aloud protocol (while verbalising their actions and thoughts as the investigator took notes). This resulted in some small changes to wording to aid clarity and a change to make one question optional (see below).

The survey was spread primarily via social media and participants were prompted to forward the study to their contacts. This resulted in 79 respondents (73% male, 27% female, age m=30). In order to understand the sample some demographic data was collected, the majority of respondents (75%) had graduated university, and 56% of respondents were employed for wages, whereas 30% were students. Full demographic information is available in Appendix B.

Though, as mentioned in Chapter 2, online news consumers tend to be young, well educated males, the proportion of male respondents in this sample is higher than would be expected on a typical online news website. This may lead to some gender bias in the results.

The participants were first asked general demographic information (age, employment status, education level, and gender) and then shown a list of technical news systems (shown in Figure 4.1). From this list, they selected those they were familiar with.

Once the systems were selected, they were presented with up to ten randomly ordered news systems with which they have familiarity.

For each system, the participants were asked to rate the system against each of Meyer's factors using a five point Likert scale, to give the reasons for these ratings, and asked how often they interact with the systems. The interface used to collect credibility ratings can be seen in Figure 4.2.

The "reasons" question was made optional when it was discovered during experiment validation that participants sometimes felt unable to answer this question and this discouraged them from continuing with the study. However, it was important that it remained as an optional question as it would provide a way of understanding the reasoning behind any patterns in the results.

The question regarding the level of interaction with the system was added to allow for grouping the responses by level of interaction. This is to ensure that the credibility levels

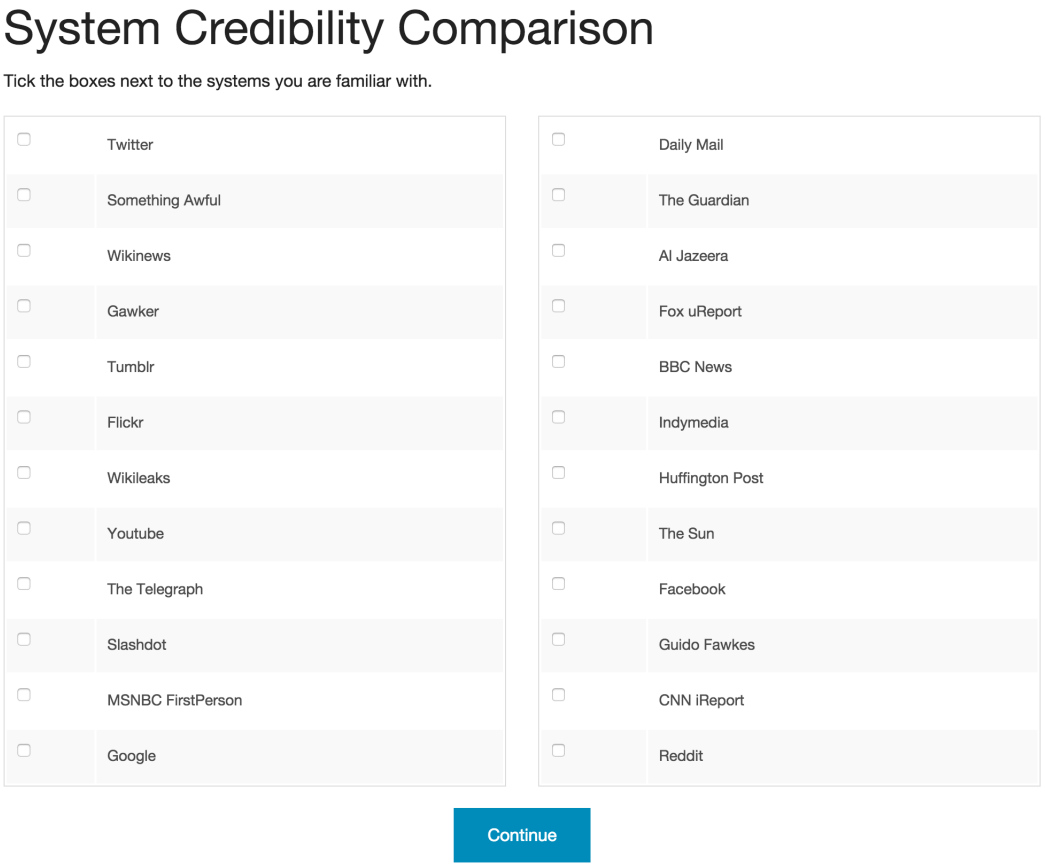


Figure 4.1: The credibility experiment system selection

are not primarily decided by respondents who have never interacted with the systems in question.

All of the systems from the spectrum in Chapter 3 were available for rating. "The Huffington Post OnTheBus" was renamed to "OnTheBus (by The Huffington Post)", "The Guardian MP Expenses" was renamed to "MP Expenses Investigation (by The Guardian)", "The Guardian Comment Is Free" was renamed to "Comment Is Free (by The Guardian)", and "Al Jazeera Your Media" was renamed to "Your Media (by Al Jazeera)" as it was found during experiment validation that several participants misunderstood the question as referring to the parent system and answered accordingly.

Context emerged as an important part of the credibility models proposed by Hilligoss and Rieh (2008), Wathen and Burkell (2002), and Fogg (2003). In this study, context is presented as "on a typical visit", which allows the participant to decide for themselves as to why they would be interacting with this news system. The focus was placed on "a typical visit" rather than on a single piece of content presented by the news system to avoid conflating message or source credibility with the media credibility being studied.

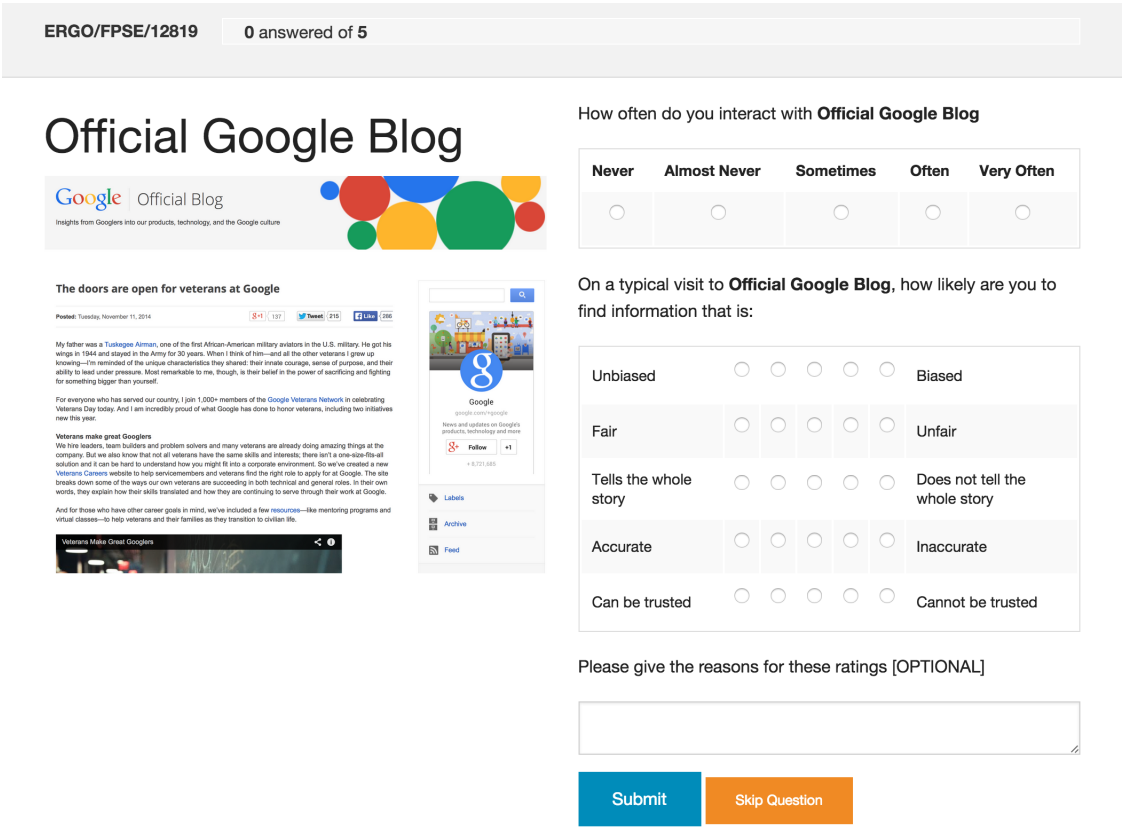


Figure 4.2: The credibility experiment question interface

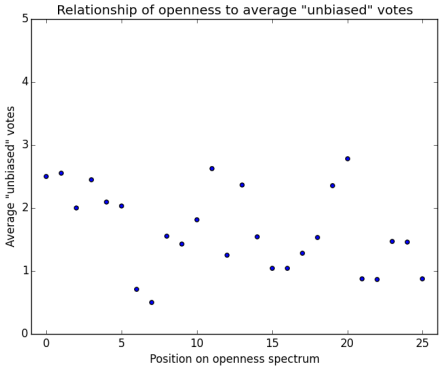
4.3 Results

Participants submitted ratings for 27 of the 32 systems presented. For each system this included a rating between 0 and 4 for each of the five factors: unbiased; fair; tells the whole story; accurate; and can be trusted, and a rating between 0 and 4 representing the level of interaction with the system (0 = never, 4 = very often). 15% of ratings were accompanied by a reason.

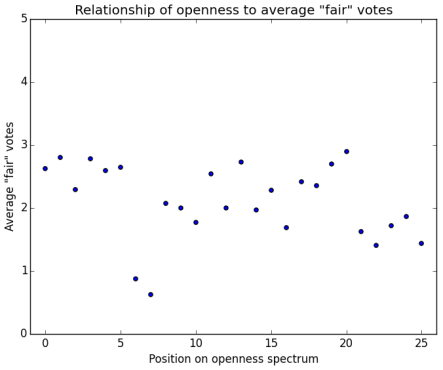
Several systems received no ratings at all: OffTheBus (by Huffington Post), Fox uReport, CNN iReport, MSNBC FirstPerson, and Your Media (by Al Jazeera). Indymedia UK received only a single rating and so no reliable statements can be made about its credibility. These have all been removed from the analysis, leaving 598 ratings in total, 26 systems rated, with an average of 23 ratings per system ($\sigma = 8.85$).

For an overview of the results, Table 4.1 shows the spectrum of credibility, based on the average of all votes received for each system. Figure 4.3 shows the relationship between the average rating for each factor and the system's position on the openness spectrum. Full response data is available in Appendix B.

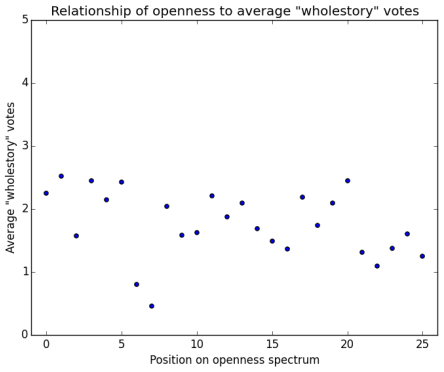
As the position on the openness spectrum and the values contributed for each factor are ordinal, Spearman's Rank Correlation will be used to determine relationships between



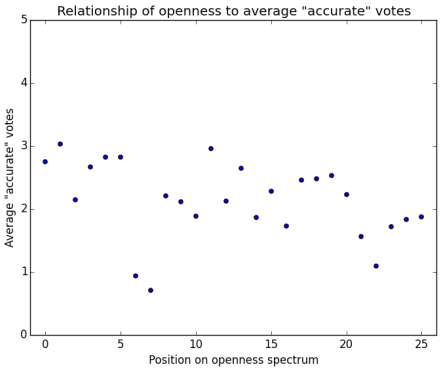
(a) Unbiased



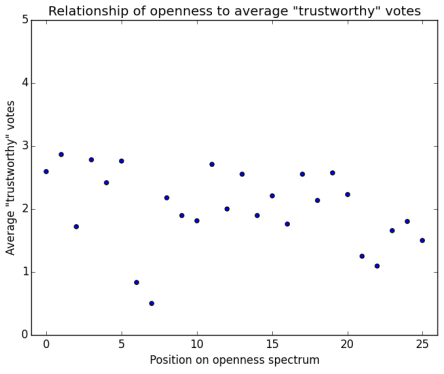
(b) Fair



(c) Tells the whole story



(d) Accurate



(e) Trustworthy

Figure 4.3: Plots of ratings of Meyer's factors against position on the openness spectrum

Credibility Position	**System**	**Openness Position**	**Change**
1	BBC News	2	-1
2	MP Expenses Investigation	4	-2
3	Wikileaks	12	-9
4	BBC News On Facebook	1	+3
5	The Guardian	6	-1
6	Wikinews	21	-15
7	Slashdot	14	-7
8	Flickr	20	-12
9	Offical Google Blog	5	+4
10	Something Awful Forum	18	-8
11	Reddit /r/AMA	19	-8
12	The Telegraph	9	+3
13	Gawker	3	+10
14	Comment Is Free (by The Guardian)	16	-2
15	Guido Fawkes	13	+2
16	Republic on Facebook	10	+6
17	Huffington Post	15	+2
18	Sky News on Twitter	11	+7
19	Youtube	25	-6
20	Twitter	24	-4
21	Tumblr	17	+4
22	Reddit /r/PoliticalDiscussion	26	-4
23	Reddit /r/Politics	22	+1
24	Facebook	23	+1
25	Daily Mail	7	+18
26	The Sun	8	+18

Table 4.1: Spectrum of credibility, and difference from spectrum of openness

the variables. The correlation between position on the spectrum and each of the factors is shown in Table 4.2. Each correlation is modest though statistically significant. Table 4.2 also shows the correlation when the biggest outlying systems (The Daily Mail and The Sun, discussed below) are removed.

Due to previous findings, such as that by Armstrong and Collins (2009), which found that increased exposure to a news media increases credibility, these correlations were also calculated when removing ratings by respondents who indicated they had "never" interacted with the system (despite previously having stated that they were familiar with them). This does not significantly change the level of correlation. All factors still correlate negatively with position on the openness spectrum (fair = -.161, unbiased = -.201, whole story = -.155, accurate = -.220, trust = -.184, ps $<$.01).

Factor	Correlation	Significance	Outliers removed	Significance
Fair	r_s = -.176	df = 598 p <.01	r_s = -.285	df = 544 p <.01
Unbiased	r_s = -.220	df = 598 p <.01	r_s = -.307	df = 544 p <.01
Whole Story	r_s = -.181	df = 598 p <.01	r_s = -.272	df = 544 p <.01
Accurate	r_s = -.236	df = 598 p <.01	r_s = -.349	df = 544 p <.01
Trustworthy	r_s = -.202	df = 598 p <.01	r_s = -.315	df = 544 p <.01

Table 4.2: Correlations between openness and the five credibility factors

4.4 Discussion

Overall, there is a modest correlation between the openness of a news system and the credibility of that system. This correlation is significant for each of the five credibility factors.

The two clear outliers in all measures are The Sun (a populist British tabloid) and The Daily Mail (a right wing British tabloid), which despite being very closed systems, were rated as bottom and second bottom overall for all credibility measures. These systems received a number of negative comments including "media coverage of the lowest grade", "sensationalist racist tripe", "so awful it's funny", and "right wing propaganda" though the Daily Mail did receive positive votes from some users, and one positive comment of "closer to the story than some of the others".

From the comments, the reasons for the low levels of credibility appear to be either that these outlets are known for overly sensational stories, or because the ideological position of the sources differ strongly from those of the participants. The fact that some respondents did visit these websites for entertainment despite finding them non-credible lends credence to the first possibility, but the comments referring to "right wing propaganda", and "racist tripe" indicate that for some it is the political position of the news source that causes the low reputation.

The first thing to note from the comments is that many participants didn't consider some of the systems to be news systems at all. Flickr got comments such as "Last time I checked, Flickr was a picture upload site, not a source of information", "I go to Flickr for photography, not news reporting", and "I don't consider Flickr a source of news", Youtube got comments including "This is mainly for entertainment, and is not reliable for more serious news", and one participant said of Tumblr that it includes "more porn than news".

Others felt that the broad platforms could not be judged as a whole. For example one participant said of Slashdot "It's information that is posted by individuals which have different levels of credibility. So there's no way to rate the entire platform as a whole!", and of Flickr "a collection of individual's uploads, and will reflect those people's views. I wouldn't expect any attempt at impartiality". Some commented that browsing habits dictate the credibility of the media received. For example Youtube received comments including "It really depends on what you watch", and "Pretty much depends on your browsing habits". Twitter received the comment "Depends on who you follow!".

The comments did re-enforce some existing research, such as the comment claiming of Youtube that "Seeing things is better than reading about them, doesn't contain personal views and perspectives" (Carter and Greenberg, 1965; Gaziano and McGrath, 1986), though others claimed that Youtube is "a soapbox for good information and total bull at the same time".

When rating the more traditional systems as credible, many participants alluded to an editorial process. For example, when rating The Guardian's Comment Is Free site, one participant said they "expect that The Guardian has a system that includes quality control" and another said that though it contains opinions, it is "edited so probably fairer and more trusted than [Reddit /r/AMA]".

The comment regarding The Guardian's "quality control" when referring to Comment Is Free provides evidence that the brand name gives users confidence in the credibility of a piece of media despite that credibility coming from a different medium. This also came through in comments for Sky News on Twitter ("Sky is a respectable news brand") and BBC News on Facebook ("generally trustworthy ... regardless of medium", "probably my most trusted brand", and "official and verified account of the BBC").

The participants held relatively positive views towards voting systems, claiming that on Reddit "User voting and high volume of users [results] in corrections" and that it is "pretty good for filtering information", and on Slashdot "User ratings help guide". One user did however comment on the potential bias of the votes ("I found the Reddit community has mostly American liberal demography, so most of their news are biased to that area") and another commented that though information is filtered by upvotes, "this doesn't necessarily make it true".

Even in the most credible news systems there were several comments referring to bias in the coverage. Of the BBC, participants said "Sometimes incomplete and a bit selective", "sympathetic to whoever is in power", and "information broadcast is down to how the organisation wants to portray a particular story", and The Guardian received comments such as "a pronounced agenda", "a very heavy 'Guardian' slant", and "have their own demographic to cater for, which means that they might be selective as to the facts they present".

4.5 Conclusions

This chapter aimed to move beyond the labels of "traditional media" and "social media" that intuitively differentiate between the sides of the openness spectrum, and find more substantive measures of the perceptions of those systems. An overview of credibility research was presented, and then a study was planned and executed which aimed to evaluate the credibility of 32 online websites.

The study found that overall there is a modest but statistically significant correlation between position on the spectrum of openness and each of the credibility factors. This correlation exists whether or not responses are accepted from participants who have never interacted with the systems, and whether or not the two outliers (The Sun, and The Daily Mail) are included.

Though there were exceptions, overall the data gathered supports the hypothesis that more open news systems are seen as less credible than more closed news systems. Looking at the comments, the reason for The Sun and The Daily Mail not following this trend is due either to their reputation for entertainment-heavy sensational stories, or due to the difference of ideological position between the systems and the participants.

Other comments showed that there exists disagreement in how the credibility of a broad platform can be rated, and even what constitutes news. Comments about systems on the more open side of the spectrum indicated that the participants did consider the outlet to be a medium that contains the contributions of many sources of differing credibility. However, when discussing the more closed systems, comments sometimes referred to the overall reputation of the company behind the system (such as The BBC and The Guardian) rather than the system itself. This is caused by the participants holding some trust in the editorial process associated with that company and indicates that they do believe that a single credibility judgement can be applied to the whole outlet. These comments were made even about The Guardian's "Comment Is Free" platform which invites contribution from non-journalists.

The comments revealed general positivity towards a voting system as a method of reducing bias, though this was not reflected in the Likert ratings, where two of the three Reddit systems were ranked in the bottom five systems in terms of overall credibility.

The correlation between credibility and position on the spectrum provides some evidence that the ordering of the spectrum presented in Chapter 3 is a meaningful representation of the variety of news systems, and indicates that credibility is an appropriate but imperfect explanation of the intuitive difference between open and closed systems.

Future work should investigate the credibility of systems that do not fit into the openness spectrum (such as liveblogs and microblogs), and investigate where these systems would

place on the spectrum of credibility. Further work is also required to investigate if this relationship exists in non-English-language media.

This analysis has shown that despite growing numbers of people who share and discuss news on open news systems, those systems lack the credibility of traditional news media. Responses indicate that there is disagreement over whether the more open systems should be considered to be "news" at all, and in some cases the editorial process is invoked to explain why traditional news is the more credible option. It may be that, to make open news systems more credible, there needs to be some form of editorial process added.

In the rest of this thesis I will investigate if it is possible to increase the credibility of open news systems by introducing a form of automated editorial control. This begins in the next chapter with an evaluation of one possible method of adding structure to news discussions. In Chapter 6, the success of this work will be evaluated using the five credibility factors used in this chapter. As there was only a modest correlation between credibility and position on the spectrum of openness, this work will also use semi-structured interviews in addition to credibility ratings in order to explore other factors that might influence a reader's perception of an open news system.

Chapter 5

Study on Identifying Groups in News Communities

The material in this chapter has previously been published as Scott et al. (2015b).

The previous chapter established that a relationship exists between the openness of online news systems and Meyer's five factors of credibility. Next I will attempt to develop techniques to improve the credibility of open news without reducing the openness to public contribution.

Both because of the editors' comments in interviews with Hermida et al. (2011) and Thurman (2008), and the comments given by participants in the study in Chapter 4, I believe that the role of the editor is an important factor in a credible news system. As identified in Chapter 2, the key responsibilities of a news editor include:

1. Attempting to stimulate debate amongst readers
2. Editing to ensure good spelling and grammar, and ensuring decency
3. Synthesizing and packaging content to share the most important news
4. Ensuring objectivity and balance
5. Avoiding duplication

This chapter will investigate techniques that can alleviate some of the issues experienced with open news systems and fulfil some of these responsibilities. This is in the expectation that better fulfiling these responsibilities will lend credibility to social news systems.

One issue which has been noted on the more open systems (towards the right of the spectrum) is groupthink (Mills, 2011). Jokingly referred to as the "hive mind" within Reddit communities, this refers to the common phenomenon whereby members of a group tend to reach a consensus and avoid dissenting opinions. Muchnik et al. (2013) performed an experiment on an un-named social news aggregation website and found that both prior positive votes and prior negative votes increased the number of future positive votes, and Mills (2011) found that on Reddit, minority opinions are "slightly marginalised but not excluded".

As receiving a response to posts made is a key determinant in the decision to continue contributing to a community (Arguello et al., 2006), this marginalisation of minority voices, especially when coupled with Reddit's interface which makes less highly rated comments less visible, could lead to discouraging future contributions from those who do not agree with the majority viewpoints, further exacerbating the problem. This is an example of what Pariser (2011) calls a filter bubble, where algorithms serve to re-enforce the existing views of internet users.

Easing this issue would potentially contribute to fulfilling three of the five editor responsibilities: having a wider range of viewpoints made prominent may help stimulate debate, increase the balance of the overall story, and avoid presenting the same viewpoint repeatedly.

As a first step towards representing multiple viewpoints within a piece of news, and ensuring that particular viewpoints are not over-represented or under-represented, it is important that we are able to recognise and group together users expressing the same sentiment towards the topic. This chapter will evaluate algorithms for grouping contribution according to sentiment, and plan and execute an experiment to evaluate the effectiveness of the chosen algorithm for grouping together content of similar sentiment in online news discussions.

In computer science, the term "sentiment" has come to be strongly associated with sentiment analysis, where algorithms are used to determine how positive or negative a piece of writing is about a given topic. When referring to "sentiment" in this chapter, I am using the common English usage meaning of "A view or opinion that is held or expressed". [1]

However, sentiment analysis would present a potential method of identifying groups of users who share a viewpoint in news discussions. Sentiment analysis is "the task of finding the opinions of authors about specific entities" (Feldman, 2013), and is typically solved either using machine learning techniques or by determining the semantic orientation of specific terms (e.g. looking for words such as "excellent" and "poor"). Sentiment analysis techniques have difficulty with language that includes spelling or grammatical

[1] According to Oxford Dictionaries `http://www.oxforddictionaries.com/definition/english/sentiment` accessed March 2016

mistakes, and those which use sarcasm, though solutions to sentiment analysis of sarcastic statements have been suggested (e.g. by Tsur et al. 2010). For a review of sentiment analysis research see Feldman (2013).

Due to the free-form and unpredictable nature of online news discussion (which can include differing levels of grammar and spelling, and regularly feature sarcasm), using sentiment analysis techniques to group together content is not feasible for my purposes. An alternative method is to produce a graph of users in the discussion, and then use community detection algorithms on that graph. This requires some method of identifying the similarity of viewpoint between users in order to form the graph. For this, I look towards online voting behaviour, which will be reviewed in the next section.

5.1 Online Voting

Voting systems are a standard method of filtering and sorting user generated content on the web, and voting is in use in almost all of the systems placed on the spectrum in Chapter 3. The more traditional systems to the left of the spectrum have been slower to adopt these techniques due to "concerns about reputation, trust and legal issues." (Hermida and Thurman, 2007), though due to increasing amounts of public contribution, traditional methods of sorting and filtering are no longer sufficient.

For example, prior to October 2005 BBC News manually moderated all public contributions published on the website (Thurman, 2008), however with the increasing amounts of participation this became infeasible. Thurman (2008) gives an example of one story in 2004 where the BBC received 100,000 contributions in one week, and only a "tiny percentage" of these contributions could be vetted and published.

In October 2005 the BBC changed their policies to allow for retroactive moderation of contributions on a small part of the website. Richard Deverell, the head of BBC News Interactive described the change to Thurman (2008) as:

> readers [can] rate each comment . . . on how interesting or useful they found it. The best ones . . . rise to the top so you [can] look at all the comments with the most highly rated ones at the top of the list or . . . look at all of them either by the name of the person who posted them or the time they were posted.

This is fairly typical of the technique commonly used across news discussion systems including those used on Reddit, The Guardian, and CNN. On some websites users are given more granular rating tools (e.g. they are asked to rate contributions on a scale, or even on multiple scales), but on most they are given the simple choice of

voting up or down, or even just of "liking" or not liking a piece of content. When discussing crowdsourcing moderation of news comments, Park et al. (2016) said that commonly used approaches have limitations, specifically that "selections don't convey an editorial voice, there is no central oversight to ensure balance, and selections may exhibit undesirable popularity biases".

Kriplean et al. (2012) view these "like" buttons as overloading two functionalities: providing a way to recognise and appreciate a speaker, but also including an implicit agreement with the content. Using the buttons in this way is contrary to the rules of many popular voting websites[2] but as mentioned, Reddit users regularly discuss the "hive mind" and it has been shown that there is at least some marginalisation of minority views in social news (Mills, 2011). Additional evidence that users vote in this way is offered by the existence of "voting brigades", where a group of aligned users vote for a piece of content en-masse to promote or silence a particular viewpoint (for example see Reddit[3] and Digg[4]). It's possible that personalities and relationships could also influence voting behaviour, with people voting for friends and allies or against enemies. However, when Cheng et al. (2014) investigated four "large comment-based news communities" though they found a number of back-and-forth arguments (with each side voting against the other), they also found that users "do not usually create enemies that continue to follow them across threads and down-vote any posts they write".

In cases where voting brigades occur, where personalities influence voting, or simply where the votes are used to indicate agreement rather than to support high quality content, the sentiment of users towards the topic should show itself in the voting behaviour. It is only in cases where the rules regarding voting are strictly followed that there should be no relationship between votes and sentiment.

5.2 Community Detection

It is possible to use these votes to produce a graph of the users of a news discussion system where the edges between users represents the amount of agreement between them. This can be calculated by counting the number of times they voted the same way about a piece of content minus the number of times they voted a different way. This graph can then be used to generate communities of users who share a similar sentiment towards the topic of the article. The largest communities would represent the most prevalent distinct opinions about the topic.

[2] e.g. Reddit "Moderate based on quality, not opinion" from `http://www.reddit.com/wiki/reddiquette` or Slashdot "simply disagreeing with a comment is not a valid reason to mark it down" from `http://slashdot.org/faq`

[3] `http://www.dailydot.com/news/week-reddit-downvote-brigades-business/` The Daily Dot - Upvote: This week on Reddit November 2011 - Accessed May 2016

[4] `https://tech.slashdot.org/story/10/08/09/1349247/buried-by-the-brigade-at-digg` "Buried By The Brigade At Digg" August 2010 - Accessed May 2016

Community detection techniques are commonly used for analysing real-world networks such as community organisations and scientific collaborations (Fortunato, 2010). They have also been used to improve friend recommendation systems and collaborative filtering techniques. Some have had success using both sentiment analysis and graph-based community detection techniques within a single system (e.g. Jaffali et al. 2014; Parau et al. 2013) but the complexity of these techniques can increase the difficulty of implementation and the time taken to evaluate a network.

As online users are already voting for content, this data already exists in an accessible way, and techniques for extracting communities from networks are well established, graph-based community detection is the method that will used in this chapter to detect sentiment communities in news discussions. Voting that they "like" or "dislike" content is one of the more common interactions users take in online communities, including Reddit, Facebook (Hampton et al., 2011), and Twitter (Meier et al., 2014), and, as discussed earlier in this chapter, it represents a user's viewpoint towards a piece of content.

Once this like and dislike data is turned into a network of users with edges representing associations (common likes and dislikes), the network can be partitioned into groups of users who share sentiment using one of the many available community detection algorithms.

Community detection algorithms can be roughly separated into three categories: divisive algorithms, agglomerative algorithms, and optimisation algorithms (Blondel et al., 2008). Divisive algorithms, such as those presented by Girvan and Newman (2002), Newman and Girvan (2004), and Radicchi et al. (2004) detect and remove inter-community links, agglomerative algorithms such as the one by Pons and Latapy (2006) recursively merge communities, and optimisation algorithms such as those by Clauset et al. (2004) and Blondel et al. (2008) attempt to maximise some function.

Community detection algorithms are typically evaluated based on the modularity of the communities detected. Modularity is a value between -1 and 1 that represents the density of links within communities compared to the density of links between communities.

Optimising for modularity is a popular approach to graph-based community detection, though Fortunato and Barthelemy (2007) identified a problem with these techniques, showing that modularity "contains an intrinsic scale that depends on the total number of links in the network", and that communities which are smaller than this scale may not be detected at all. This is particularly a problem in very large networks where smaller communities will not be detected.

The method used in this chapter is the Louvain Community Detection algorithm, as proposed by Blondel et al. (2008). It is a heuristic based method which optimises for modularity. It is simple to implement, performs well (achieving a high level of modularity

in low computing time), and runs well even on large networks. It also partially deals with the resolution limit problem identified by Fortunato and Barthelemy. The algorithm is detailed in the next section.

5.3 Louvain Community Detection

The Lovain Community Detection algorithm was proposed by Blondel et al. in 2008. It is very fast, with linear complexity on typical data (Blondel et al., 2008). The algorithm begins with each node assigned to its own community, and repeatedly executes two phases. The first phase is described by Blondel et al. (2008) as:

> For each node i we consider the neighbours j of i and we evaluate the gain of modularity that would take place by removing i from its community and by placing it in the community of j. The node i is then placed in the community for which this gain is maximum (in case of a tie we use a breaking rule), but only if this gain is positive. If no positive gain is possible, i stays in its original community. This process is applied repeatedly and sequentially for all nodes until no further improvement can be achieved and the first phase is then complete.

The second phase involves creating a new network whose nodes are the communities created during the first phase. The weights of edges between the new nodes are the sum of the edge weights between the two communities. The first phase is then performed on this new network, and these two phases are repeated until no further changes can be made.

Due to the recursive nature of the algorithm, it provides intermediate stages which allow for different levels of granularity in the groupings. This feature helps to partially avoid the resolution limit identified by Fortunato and Barthelemy (2007) as selecting different levels will result in communities of different sizes appearing. These intermediate stages provide a hierarchy of communities which may be useful when identifying groups of users with similar sentiment.

In their paper, Blondel et al. (2008) test the algorithm's performance by running it against a small social network, a network of scientific papers and their citations, a sub-network of the internet, a network of webpages, and other datasets. It performs well in all cases (having a high level of modularity with a small computation time). To verify that the Lovain Community Detection algorithm will be able to group together news discussion participants who share similar sentiment, this study involves generating a graph using news comments and employing a web-based study (ERGO #10814) to explore the accuracy of the groupings.

5.4 Methodology

I looked to the comment systems of the websites analysed in Chapter 3 for systems which 1) have enough activity to generate a large amount of data and 2) have APIs which allow access to likes/dislikes. Four of the systems provided the data needed to create a network for use with the Louvain method: Facebook, Twitter, CNN, The Telegraph.

Of these, The Telegraph and CNN share a commenting system (Disqus), and therefore have a common API. They also, as dedicated news websites, have a concept of a "story" whereas the other two systems do not. This makes it easier to group together users based on sentiment without needing to first separate users based on topic. For these reasons, this experiment used data from The Telegraph and CNN websites.

5.4.1 Data Collection

I collected every news story which, between 22/05/2014 and 27/05/2014, featured in the "Most Popular" section of CNN or the "Most Viewed" section of The Telegraph. This resulted in 45 stories: 12 from CNN and 33 from The Telegraph. The difference can be attributed to the fact that The Telegraph's "Most Viewed" section shows ten items at a time, whereas CNN's "Most Popular" section shows only six.

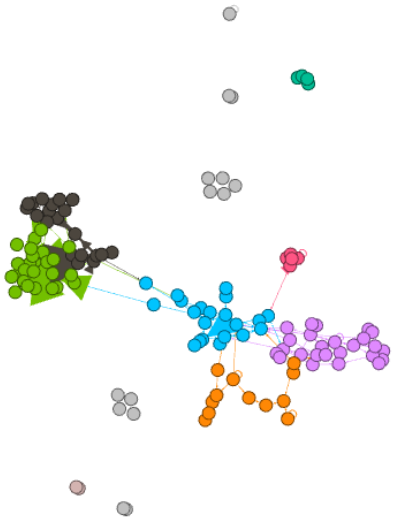
All comments on each of these stories were gathered, including which users had "liked" and "disliked" which comments. The resulted in 23,655 comments, 43,658 likes, and 5,828 dislikes.

From this data, for each story, a weighted graph was produced of relationships between the people who interacted with that story. This gave an average of 202 nodes ($\sigma = 264$) and 822 edges ($\sigma = 1626$) per graph.

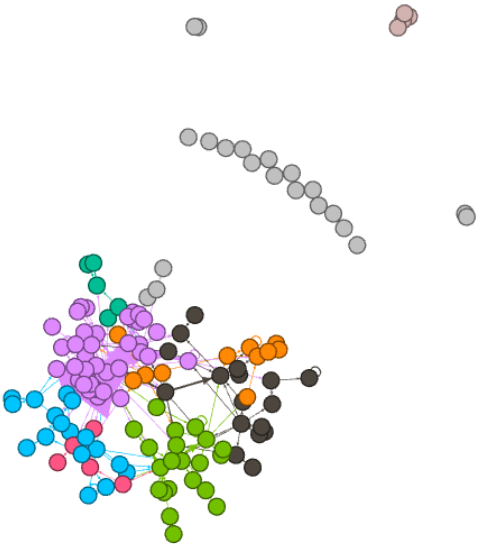
The weight of the edge between nodes i and j is given as the total number of instances where $ci_n = cj_n$ minus the total number of instances where $ci_n \neq cj_n$ for each comment n where ci and cj are all the votes by users i and j.

$$[i, j] = \sum_{n=1}^{N} \begin{cases} 1 & ci_n = cj_n \\ -1 & ci_n \neq cj_n \end{cases} \tag{5.1}$$

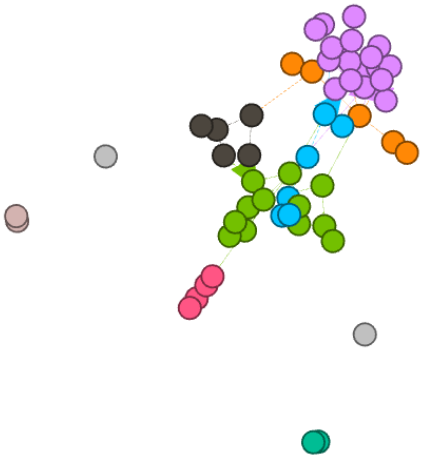
When used on these networks, the Louvain algorithm generates a total of 877 groups (an average of 19 groups per story, $\sigma = 24$) with an average of 10 members per group ($\sigma = 29$). Some examples of the communities detected are shown in Figure 5.1.



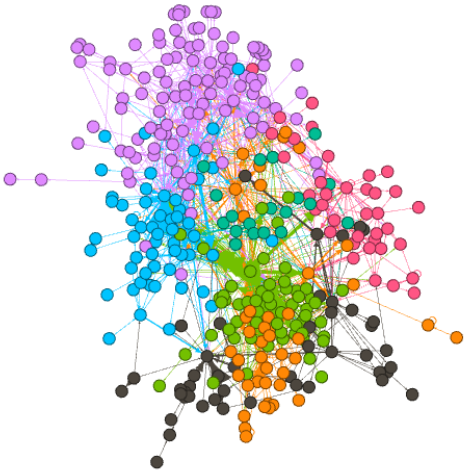
(a) "Scientists: U.S. historic landmarks imperiled by climate change" from CNN. 127 nodes, 179 edges, 20 groups.



(b) "Thailand wakes to military rule: What it means" from CNN. 139 nodes, 210 edges, 25 groups.



(c) "How ITV missed the best moment of the FA Cup final" from The Telegraph. 55 nodes, 86 edges, 10 groups.



(d) "Europe's centre crumbles as Socialists immolate themselves on altar of EMU" from The Telegraph. 367 nodes, 1225 edges, 11 groups.

Figure 5.1: Detected communities for four stories

5.4.2 Study

To determine how well the algorithm performed, this study will investigate how often pairs of comments from users placed in the same group are considered to be similar, when compared to pairs of comments from users in different groups. The hypothesis to be tested is that a pair of comments from users in the same group will be voted as "similar" more often than a pair of comments from users in different groups.

To generate the questions for this study, every pair of comments posted in reply to a common parent was identified. These pairs were separated into two groups: those pairs where both comments were by users from the same group, and those where comments were by users from different groups. The different-group pairs were then filtered to ensure that for each parent comment, there was only one pair for each combination of groups, and the same-group pairs were filtered to ensure that for each parent comment there was only one same-group pair per group. After this filtering, 3554 pairs remain, 1658 with both authors from the same group, and 1896 with authors from different groups. To make this manageable to test, fifty pairs with same group authors and fifty pairs with different group authors were randomly selected. This gave 56 pairs from The Telegraph and 44 pairs from CNN.

Participants were then invited to view these comment pairs through an online interface. The study was shared through social media and participants were invited to share the study with their own contacts. This resulted in 154 respondents.

The participants were presented with twenty pairs of comments and asked to decide if they felt the pairs of comments shared a similar sentiment regarding the parent comment. The interface for rating the similarity of sentiment can be seen in Figure 5.2.

5.5 Results

Of the 154 respondents, 20 submitted fewer than 10 ratings (giving up quite early in the process) and so their results have been removed. This leaves 2376 ratings from 134 participants, 1098 votes for "similar", 838 votes for "dissimilar", and 404 votes for "can't tell". The "can't tell" votes will be treated as an inability to answer, which leaves 1936 usable votes. The distribution of the ratings can be seen in Figure 5.4, an example question can be seen in Table 5.1, and the full record of responses is available in Appendix C.

For each question, the percentage of respondents who said the comments were "similar" was calculated. As the data are not normally distributed, the Mann–Whitney test was used to compare the two groups (see Figure 5.3).

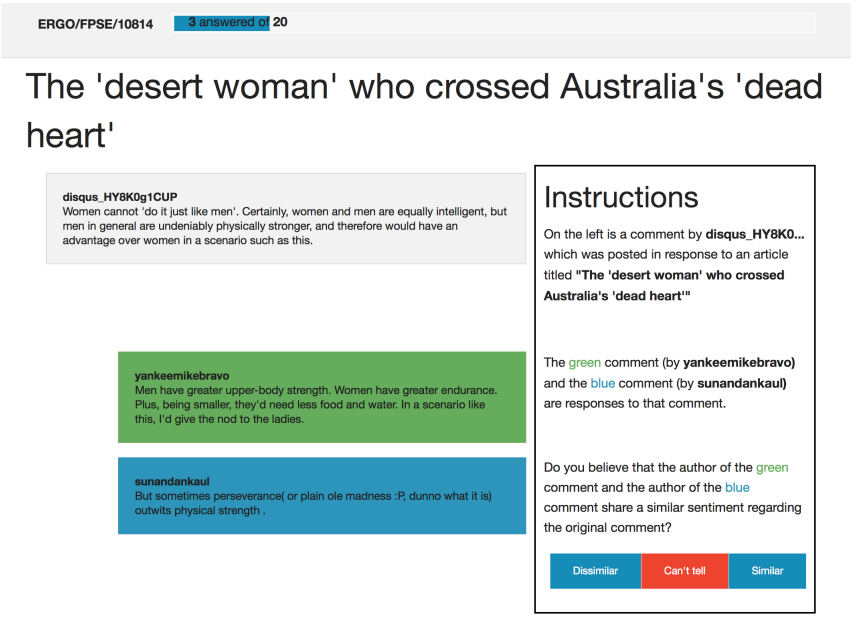


Figure 5.2: The sentiment experiment interface

On average, comments by authors who were allocated to the same group had a higher percentage of "similar" answers (M = 66.91) than comments by authors allocated to different groups (M = 43.6). This difference is significant (U = 653, p <.001).

I used the grouping status (same group or different groups) as a vote by the system (for similar and dissimilar respectively), and calculated how often the system's vote agreed with the votes of participants (62%). I also calculated this information for each participant, which is plotted on a histogram in Figure 5.5. Also plotted is the level of agreement expected given random groupings (45%), the participants' overall average level of agreement (66%), and the standard deviation (9).

5.6 Discussion

The significant difference between the percentage of "similar" votes in the same group questions and different group questions indicates that using the Louvain Community Detection algorithm on a network of votes allows for identifying communities of users who share sentiment. This suggests that the assumption stated at the beginning of this chapter, that users vote for content they agree with rather than that which is high quality, is at least somewhat true.

To find how meaningful this difference is, I calculated how often the system agrees with each participant when voting, and also how often participants agree with each other. The average agreement of the system (62%) is within a single standard deviation (9) of the participants' average level of agreement (66%).

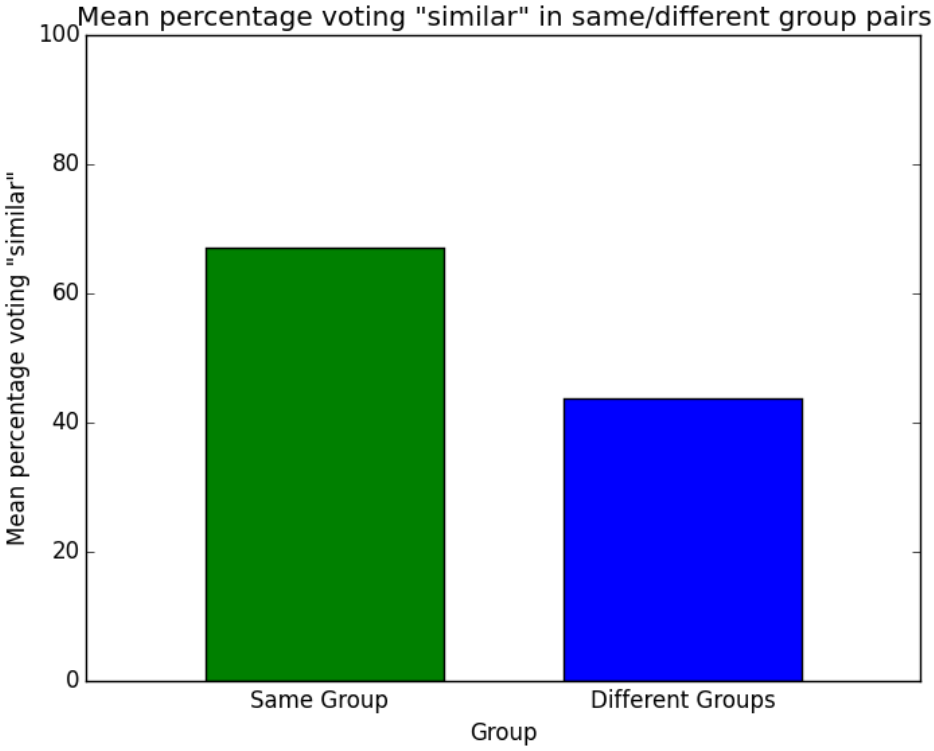


Figure 5.3: The mean percentage of participants voting "similar" in same-group and different-group pairs

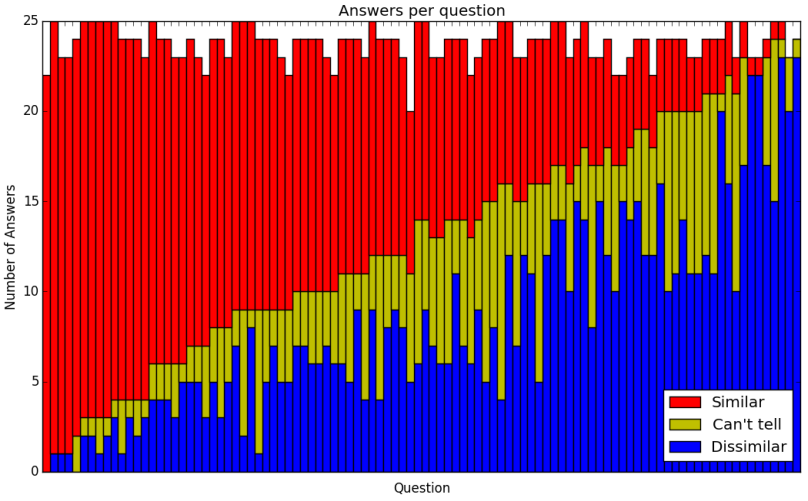


Figure 5.4: The distribution of sentiment ratings

Question ID	36682
Groups	Different
Parent Question	**AlisonHell**: If the King did not approve the coup that kick Thaksin out, but let democracy does its job back then, Thailand would not be divided as much as today. where is the King now? It's time for him to clean up the mess he and his people created. "I believe Thai both side still listen to him". it is not too late for him to make it right. He need to say and do something now. the longer he wait the worse it will get for his country and his people. And he can only blame himself for the problem he took part in creating it.
Comment 1	**geoffmaguire**: The King and Queen are extremely old, impaired and live in seclusion in a hospital. He only makes the occasional public appearance on his birthday etc and no longer has the capacity to influence events.Their son and heir apparent is in the military so...
Comment 2	**disqus_P3aOImzQzv**: the scenario is simple: The king or his son ordered the general to move... they are going to twist the arm of the acting Pm to resign and appoint an interim party.. Then the army will descend on the Red and kill 100 or 200 on the excuse is that they have started something. some crazy excuse to satisfy the US three years later, they will have a controlled election.. and that is the end of that...Thailand has no democracy and never will SO the part I do not understand is that why the Japanese are still doing business there or the Germans, or the Koreans.. it is a corrupted infested state.
Similar	5
Dissimilar	15
Can't Tell	4
Percentage Similar	25%

Table 5.1: Example question from sentiment study

This indicates that the system is almost as successful as an average human at comparing the sentiment of two comments from a single news discussion. Both of these agreement levels are much higher than that which would be expected given random groupings (45%).

To further evaluate the quality of the sentiment groupings would require access to existing news editors who would be willing to submit their own votes on the pairs selected

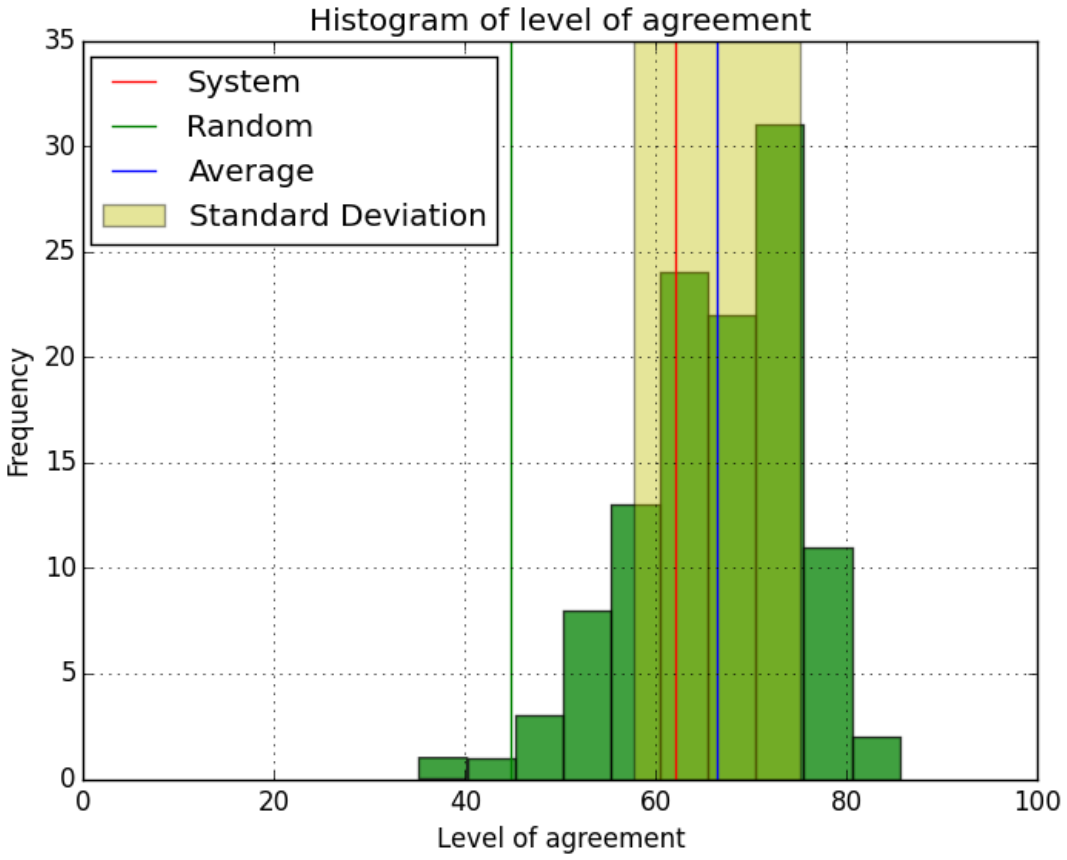


Figure 5.5: The level of agreement between participants

for this study. This would allow a comparison between the use of the algorithm and the current best practice.

The relatively low levels of average agreement may indicate that there is significant disagreement over the level of subtlety of opinion which is considered "dissimilar sentiment". This may be related to the prior knowledge of the reader, who is more able to see subtle differences in positions similar to their own, and more motivated to closely scrutinise the messages (Petty and Cacioppo, 1986; Wathen and Burkell, 2002) than when the messages represent positions significantly different to their own. This should be investigated in future work and could lead to the development of systems which use the views of the user to present a broad range of viewpoints, rather than presenting a single set of content for all users, or, as in the case of most "personalised" news systems, only presenting viewpoints which the user agrees with.

5.7 Conclusions

This chapter proposed a method of reducing the issues of groupthink prevalent in open news systems by utilising a community detection algorithm to group content based on

sentiment. It was suggested that this would help fulfil three of the roles news editors traditionally play. The success of this technique relied on an assumption that news discussion participants would not strictly obey the instructions to "vote based on quality" and would instead allow their biases to show through their voting behaviour.

After identifying a suitable algorithm, a study was issued which gathered 1936 votes for the similarity of pairs of comments. The groupings were used as an indicator of if the comments should have similar or dissimilar sentiment, and it was found that there was a significant relationship between a pair's grouping status (same group or different groups) and the percentage of respondents voting that the pair shared similar sentiment. The system was then compared to each participant, and it was found that its level of agreement was almost as high as the average level of agreement between participants.

This validates the assumption about the way that news discussion participants vote, though due to the relatively low level of agreement amongst participants further research is required into ways of presenting information according to individual viewpoints rather than a single view for all users. Future work may also attempt to improve the grouping algorithm, revisiting the possibility of adding sentiment analysis to the community detection algorithm as Jaffali et al. (2014) and Parau et al. (2013) did, or investigating the other promising community detection algorithms, which work in different ways and so can have better or worse results in different domains.

However, with the performance and accuracy of the algorithm confirmed, the next stage in this work is to examine if presenting this extra structure as part of the discussion makes a noticeable change to the way that people interpret and interact with the content. This is the content of the next chapter.

Chapter 6

Study of the Effects of Novel Comment Presentation on Reading Behaviour in Online News

Previous chapters identified a difference in the credibility of online news outlets relating to their openness to public contribution, and found that these differences may be partially explained by the lack of editorial control in the more open systems. After identifying a potential solution to this, of using community detection algorithms to identify related content and present a range of viewpoints, I showed that the chosen algorithm, the Lovain Community Detection algorithm, is able to group together related comments almost as well as an average human, creating additional structure not present in existing news systems.

This chapter will execute a study to investigate the effect of this additional structure on the way that people read news discussion. The study will be conducted by exposing groups of users to different combinations of algorithms and interfaces, and measuring the topics they identify, their perceived credibility, and their usability ratings of the system. For further insight into the reasons for the credibility ratings, a number of participants will also be interviewed.

This chapter will investigate the final research question: Will promoting a wide range of viewpoints (using viewpoint groupings) change the exposure of readers to different viewpoints, and alter the perceived credibility of a system? The hypotheses investigated in this chapter are:

1. Adding structure using the technique detailed in Chapter 5 will result in an increase in credibility.

2. Adding this structure will result in an increase in the number of points of view encountered during reading.

6.1 News Comments

This study will investigate how people respond to news comments, rather than to primary news content. There are disadvantages to this, in that news comments are not the same as news content: they have been described as being "less thoughtful and more impulsive [than other forms of audience participation]" (Reich, 2011), they tend to contain more opinions and fewer facts, and do not have the same goal of unbiased/neutral reporting. However, several of the websites listed in the spectrum from Chapter 3 are similar in structure to news comments: in particular Reddit, Twitter, and Facebook have very similar structure, with short opinion-heavy messages the norm.

Applying the Louvain Community Detection algorithm directly to news content would be preferable. However, this would require a source of news content with short contributions by multiple authors, so that the authors could be grouped by sentiment, and the content presented in context. Live Blogs, such as those ran regularly by The Guardian, are written by multiple authors but typically by one at a time, are written in a single voice, and present a single narrative. Contributions on social news websites such as Reddit do fit the criteria but due to their anti-spam methods[1], information regarding which users vote for which content is not available. Appendix K details plans for possible future work that creates such a data set.

Basing this experiment on news discussions rather than news stories has a number of advantages. It allows for using existing designs that are currently in use in online news communities rather than requiring wholesale redesign of a new system. It also means that there is content that fulfils the requirements of the study available to access on existing online news communities.

6.2 Interface Design

With access to previously unavailable structure (in the form of the clusters of contributors), a decision must be taken regarding how best to present this new information within news discussions.

One option is to present the discussions in the same way that they are presented currently, with the only change being the way that the contributions are ordered. This would allow for confidence that any behavioural changes are solely due to the use of

[1] "Please note that the vote numbers are not 'real' numbers, they have been 'fuzzed' to prevent spam bots etc." According to the Reddit FAQ `https://www.reddit.com/wiki/faq` accessed November 2016

the community detection algorithm, and not to a change in interface. However, existing news discussion interfaces were not designed with this structure in mind and as such may not take full advantage. This may reduce any impact on behaviour, making it difficult to measure. Also, previous experiences with similar looking systems may influence interactions with and judgements of the designed system.

An alternative would be to design an entirely new interface built around this extra structure. This could best present the structure free from the constraints and expectations of the existing design, but it would be difficult to separate changes that are due to algorithm from changes due to interface design.

To best separate the effects of algorithm change from interface change, an experiment will be performed in which participants are placed into one of four groups. One group will be presented with the existing design and algorithm, one with the new design and existing algorithm, one with the existing design and new algorithm, and one with both new design and new algorithm.

It has been shown (e.g. by Metzger et al. 2003; Fogg and Tseng 1999; Olaisen 1990) that design features and quality can affect the credibility of a system, so this experiment will need to take this into account. As the new design will be produced in a shorter time frame than existing designs (which have had many years of development and design), I will implement both the new and old interface metaphor, using the same look and feel, in order to minimise any perceived differences in clarity or quality.

This section will review appropriate literature to generate requirements for the interface to be used in this experiment, will review previous attempts at alternative discussion interfaces, and will refine and choose a final interface through a participatory design process. A full description of the design process can be seen in Appendix D.

6.2.1 Motivations

To guide the design of the new interface, this section will review literature regarding the motivations of users to read and contribute to online news and news discussions. Accounting for these motivations during the design process will increase the likelihood of participants engaging fully with the system.

In 2006, Flavián and Gurrea proposed that people read newspapers for three reasons: to search for specific information, to search for updated/breaking news, and for leisure/entertainment (Flavián and Gurrea, 2006). In 2015, 48% of UK adults said that they followed news to "find out what's going on in the world" and to "know what's going on across the UK". 43% cited knowledge about events in their local area and events going on in the country, and 32% said they "feel it's important to keep informed about certain issues" (Ofcom, 2015). For news discussions, Diakopoulos and Naaman (2011) found

from their survey of 390 users of SacBee.com (the online presence of the Sacramento Bee newspaper) that the primary motives for users to read news discussions included gaining more information on a story, including story updates; validating/comparing their own opinion; seeking entertainment; and gauging perspectives from the community.

In terms of interacting with news, Springer et al. (2015) suggest that users post comments for one of two reasons: either to publish their own opinion or to react to the article or others' comments. They also note that non-contributing users ("lurkers") can use the contributed comments to become aware of perspectives and help form their own opinions. Chung (2008) showed that politically engaged users were most likely to use interactive features on news websites, suggesting that commenting may be driven by a will for democratic participation (Springer et al., 2015) but also suggesting that current systems do not encourage participation from those already marginalised. Cheng et al. (2014) found that users who have content "disliked" become more likely to make low quality contributions in the future, and authors who received no social feedback (likes & dislikes) are most likely to leave a community.

In Sunstein's 2002 book "republic.com 2.0" (and its predecessor, "republic.com"), Sunstein warns that the freedom and choice of sources offered by new technology may not lead to better informed citizens, but instead could lead to a more polarised society as people restrict themselves to consuming agreeable information. Garrett (2009) found that in political blog comments only 13% of comments expressed disagreement, suggesting a high degree of homophily in online news communities.

If it is true that people seek only information agreeable with them, this can lead to news systems where not only are there demographic imbalances due to the medium (see Section 2.9), but also issues of overrepresentation of particular viewpoints leading to "cyberpolarization"(Sunstein, 2002b). This is exacerbated by the fact that many online news systems that support reader contribution bury unpopular content (see Chapter 3), leading to less interaction with it, and eventually to people with underrepresented opinions leaving the community (Cheng et al., 2014).

Munson and Resnick listed three problems with people consuming only information that is agreeable to them:

1. It has been shown that interacting with only agreeable information can lead to increased polarisation and to participants adopting more extreme positions (Sunstein, 2002a)
2. Exposure to diverse opinions can improve problem solving and decision making (Nemeth, 1986; Nemeth and Rogers, 1996)
3. People tend to overestimate the popularity of their own viewpoints (Sanders and Mullen, 1983)

Munson and Resnick (2010) and Freelon (2015) found that research exists both to suggest that people seek out information they agree with while avoiding dissenting information (Frey, 1986; Mutz, 2001; Iyengar and Hahn, 2009), and to suggests the opposite, that individuals seek diverse news and information (Stromer-Galley, 2003; Horrigan et al., 2004; Gentzkow and Shapiro, 2010; Kobayashi and Ikeda, 2009). Freelon (2015) explains this by referencing studies that show that when given a choice, participants usually select agreeable content "without systematically avoiding opinion-challenging content". Ludford et al. (2004) found that users were more likely to participate in discussions with diverse groups of users, especially when they were informed of their unique position or viewpoint within the group.

The "Political News Sharing" model proposed by An et al. (2014) attempts to predict a person's likelihood of sharing a political news article. The model is built on four major factors: gratification (satisfying desires such as entertainment), selective exposure (avoiding conflicting information), socialisation (someone would like to read this news), and trust & intimacy (the source is credible). They found that 90% of participants in their study subscribed to news of only one political leaning, but 41% of users were exposed to diverse news due to the sharing behaviour of their friends. They also found that people were more likely to share news from a traditional media source than news shared by their contacts, except for political news stories which people preferred to be from their friends (An et al., 2014).

To maximise the motivation to read and contribute to news discussions, when working through the design process the following points will be considered:

1. News should be informative (Ofcom, 2015; Flavián and Gurrea, 2006; Diakopoulos and Naaman, 2011), up-to-date, and entertaining (Flavián and Gurrea, 2006)
2. Readers should be able to compare the perspectives of others to their own (Diakopoulos and Naaman, 2011; Springer et al., 2015)
3. Readers should be presented with diverse content (Ludford et al., 2004; Munson and Resnick, 2010; Sunstein, 2002a)
4. Readers should have an opportunity to respond to the original author and to commenters (Springer et al., 2015)
5. Commenters should be able to identify their unique contribution within the community (Ludford et al., 2004)
6. Comments should be engaged with to encourage future contributions (Cheng et al., 2014)

6.2.2 Existing Work

There are a number of previous attempts at redesigning discussion forums to better facilitate sensemaking, to encourage contribution, and to encourage interaction between different communities of users.

In 2004, Ludford et al. performed an intervention with the MovieLens film recommender system. They added a discussion system with a focus on investigating the effects of similarity and uniqueness on motivation to participate. They found that people contributed more when they were participating in discussions with people dissimilar to themselves, and that people contributed more when their unique contribution to the discussion was highlighted to them (Ludford et al., 2004).

Faridani et al.'s "Opinion Spaces" plots 2D graphs based on commenters' self-provided "opinion profiles" (Faridani et al., 2010). This involves commenters indicating their level of agreement with five statements relating to U.S domestic politics, and then projecting these five points onto a two dimensional interface where members with high agreement are plotted close to each other and members with low agreement are more distant. Users can then navigate the discussion by clicking on the points in 2D space. Faridani et al. (2010) found that Opinion Spaces increased the level of engagement of readers, and the level of agreement and respect that readers had for comments they read, though it did not result in readers viewing more diverse comments.

Other work has explored news visualisation without the focus on diversification. These include ConVisIT, proposed by Hoque and Carenini (2015), which involves the use of textual analysis and information visualisation techniques to present readers with tools for navigating topics in a discussion; ForumReader by Dave et al. (2004) which experimented with novel visualisations and analytics to help readers navigate forum threads; and systems such as CommentIQ (Park et al., 2016) and Arkose (Nam and Ackerman, 2007) which use novel user interfaces and tools to aid human moderation of comments.

Due to the use of a community detection algorithm, the system produced in this chapter is able to indicate when contributions are similar or dissimilar to each other in a similar way to Faridani et al. (2010). It is hoped that this will result in a similar increase in engagement and respect, and also that ordering the comments as to maximise the number of groups viewed will result in interaction with more diverse content. The method advocated in this chapter avoids the need for users to provide their own values, as these are inferred from existing social behaviour. The users also do not need to learn any formal schemas or classification systems. The participants' place within the groups can be automatically communicated to them, and this may serve the same purpose as Ludford et al.'s highlighting of individual contribution, which increased contribution in their study.

6.2.3 Participatory Design

To minimise the chances of usability issues impacting on the results of the experiment, for this study the system is designed using the principles and practices of participatory design. Participatory Design is a set of techniques, tools, and ideas that emphasise the involvement of end users in the design of systems. It represents an approach towards computer systems design "in which the people destined the use the system play a critical role in designing it" (Schuler and Namioka, 1993). It is similar to user-centered design except that instead of designing on behalf of users, the designers must work with the active participation of users (Iivari, 2004). Spinuzzi (2005) describes Participatory Design as:

> It attempts to examine the tacit, invisible aspects of human activity; assumes that these aspects can be productively and ethically examined through design partnerships with participants, partnerships in which researcher-designers and participants cooperatively design artefacts, workflow, and work environments; and argues that this partnership must be conducted iteratively so that researcher-designers and participants can develop and refine their understanding of the activity.

Participatory Design grew out of Scandinavian workplace democracy and a partnership between academics and trade unions. The aim was to allow workers to influence the technologies that were changing their ways of working rather than simply accepting systems developed by management (Spinuzzi, 2005).

Practitioners of participatory design view tacit knowledge as central to the design process, understanding that users can accomplish a task in a number of ways and that there isn't necessarily one best method that can be broken into discrete tasks. They also assume that the actual users are best placed to understand how to improve the tools they use, having the most experience actually interacting with those tools (Schuler and Namioka, 1993).

For an overview of participatory design techniques see Spinuzzi (2005), Muller (2003), and Schuler and Namioka (1993). For the design of the interfaces used in this work, it was ensured that the stakeholders were given opportunities to design, interact with, and change the artefacts and workflows to be used in the final design. The details of the participatory design process used can be seen in Appendix D.

6.2.4 Final Designs

The process resulted in two completed designs, Design A is a re-implementation of existing online news discussion systems, and Design B is designed to make best use of

the new structure available. The completed designs are shown in Figure 6.1. Design A presents comments initially ordered by the aggregate number of likes, with the option to order chronologically. Each comment is associated with a small avatar and username of the commenter, the amount of time since the post was made (e.g. "20 hours ago"), and the aggregate number of likes. Replies are indented and placed below the comment they are replying to, but there is no other indication of connections between comments.

Design B first presents a representative comment for the seven largest groups, with each group assigned a unique colour. These representative comments can be clicked to filter all comments to show only those posted by members of the selected group. There are also links to remove the filter (view all groups) and to filter so it only shows posts by users who are not inside a group. Below this is a simple explanation of the filter being applied (e.g. "Showing all root posts by members of group 1 sorted by group votes"), and avatars of the users in the currently selected group.

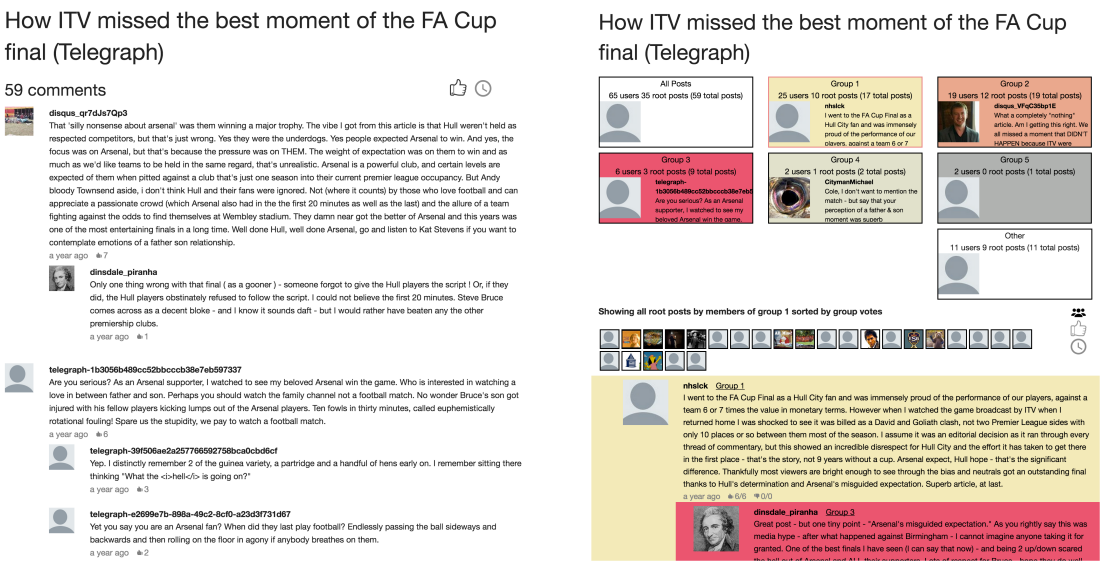
Below this is the list of comments, each comment coloured according to the group of the comment's author. Alongside the author's avatar and username is the name of the group they belong to and a link to view more posts from this group. The time that has passed since the comment was created is also displayed as is the number of likes and dislikes. On this interface likes and dislikes are split, so that it is possible to know how many the comment has overall, and how many of those were given from other members of the same group.

These comments can be sorted in one of three ways: by aggregate votes, chronologically, or by groups. Sorting by groups is the default option, and when this is selected the comments are ordered so that the first comment is the highest rated comment from the largest group, followed by the highest rated comment from the second largest group, and so on until all groups are exhausted, at which point the second highest rated comment from the largest group is shown. This method will ensure that for a given number of comments the user will be exposed to the largest possible number of groups.

6.2.5 Design Evaluation

Later in this chapter an analysis is detailed investigating the effectiveness of the interface at exposing users to different voices (members of different groups). This analysis will also include a usability evaluation of the designed system, which will ensure that any results are not simply a result of usability differences. This section documents an initial evaluation of the system to compare the number of viewpoints present when using the developed interface to those present on a typical existing news discussion system.

Previous work by Giannopoulos et al. (2014) evaluated diversification methods of comments by counting "information nuggets" present in the top 10 comments when ordered using different methods. Their work suggested that their "MAXSUM2" method is able



(a) Design A (b) Design B

Figure 6.1: Completed Designs

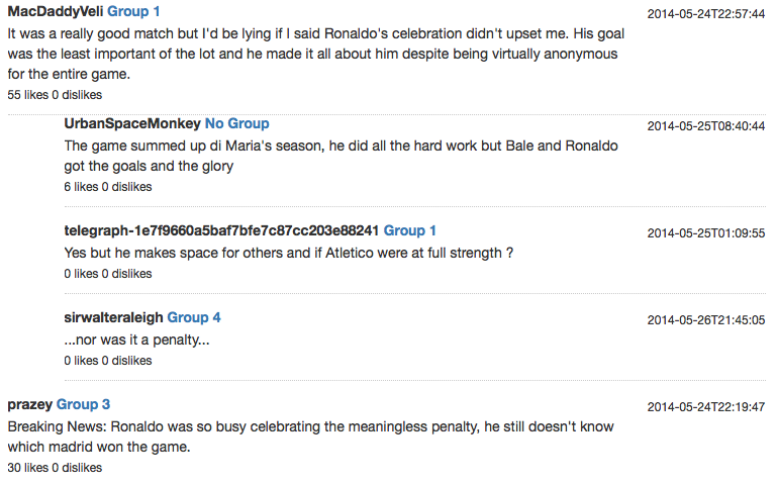
to present a greater diversity of comments than existing methods of diversifying search results, however they did not compare against the most common methods of comment organisation used in online discussion systems today (chronology and popularity).

Clarke et al. (2008) used information nuggets as a way of measuring the diversity of a collection of documents. In their work, an information nugget represents a binary property of a document (e.g. a fact, or the organisation behind the document). Since the work in Chapter 5 found that the groups generated by the Louvain Community Detection Algorithm are comparable to those generated by a human, and that the content produced by the different groups differs in sentiment, this analysis will use groups rather than information nuggets to measure the diversity in the comments in the dataset.

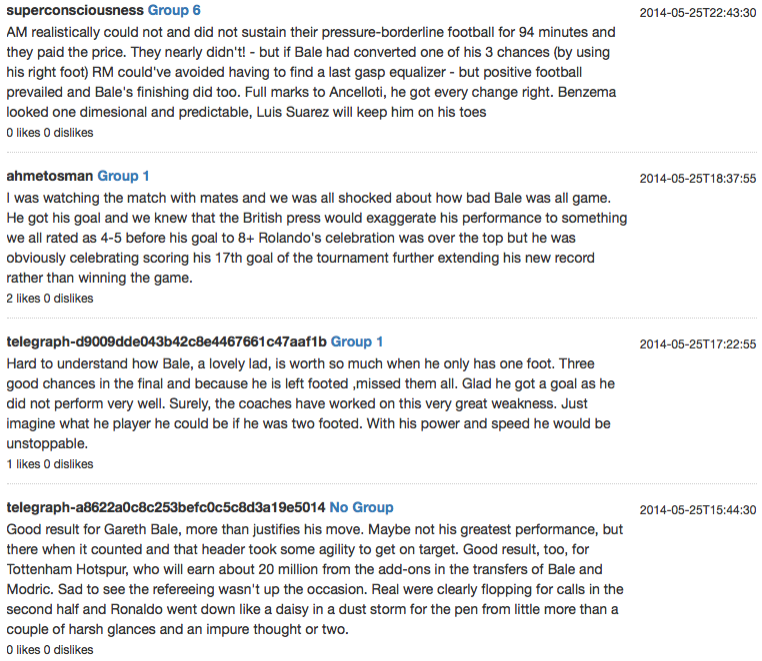
Figure 6.2 shows the top comments for the story "Bale's magic moment sees Real rule again"[2]. The top comment when ordered by popularity and when ordered by groups is the same, though the ordering of replies is different and the second root comment is different. The top comments when ordering chronologically are entirely different to the other two methods. There are 4 groups viewable in the top 5 comments when ordering by popularity, 3 when ordering chronologically, and 4 when ordering by groups. For each of the 44 discussions in the dataset, the comments were ordered according to each of the three conditions and the number of groups present in the top comments were counted. The results of this are shown in Table 6.1.

In this dataset, each discussion has on average 10.11 groups. When ordering the comments according to groups, the reader will, on average, see content from a greater number of groups than when ordering according to chronology or by votes.

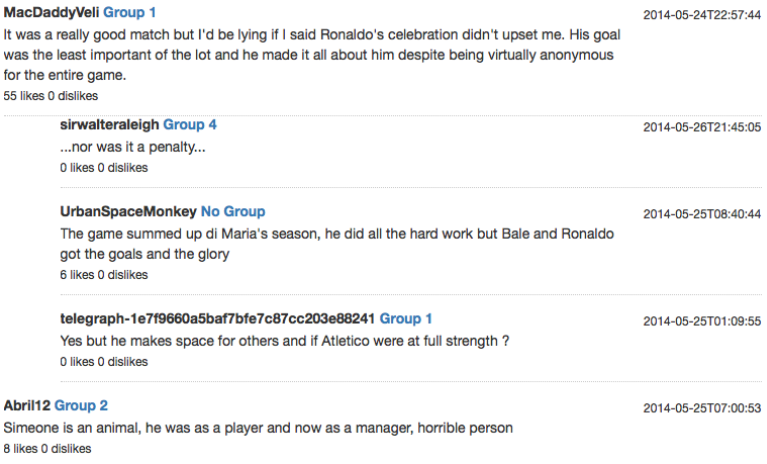
[2]Available at `http://www.telegraph.co.uk/sport/football/competitions/champions-league/10853767/Real-Madrid-4-Atletico-Madrid-1-Champions-League-final-2014-match-report.html`



(a) Ordered by popularity



(b) Ordered chronologically



(c) Ordered by alternating groups

Figure 6.2: Top comments on the story "Bale's magic moment sees Real rule again"

Comments	**New**	**Popular**	**Groups**	**Significance**
5	2.48 (σ=0.81)	2.57 (σ=0.89)	**3.39 (σ=0.88)**	H = 22.35 (p <0.001)
10	3.43 (σ=1.12)	3.89 (σ=1.35)	**4.70 (σ=1.29)**	H = 19.18 (p <0.001)
20	4.82 (σ=1.50)	4.95 (σ=1.59)	**6.05 (σ=1.69)**	H = 12.22 (p <0.01)
30	5.66 (σ=1.58)	5.61 (σ=1.58)	**6.64 (σ=1.98)**	H = 7.72 (p <0.05)
40	6.25 (σ=1.76)	6.02 (σ=1.86)	**6.98 (σ=1.98)**	H = 5.83 (p >0.05)
50	6.64 (σ=1.76)	6.41 (σ=1.96)	**7.34 (σ=2.06)**	H = 5.40 (p >0.05)
100	7.77 (σ=2.12)	7.55 (σ=2.19)	**8.30 (σ=2.26)**	H = 2.86 (p >0.05)
500	9.20 (σ=3.48)	9.27 (σ=3.70)	**9.45 (σ=3.82)**	H = 0.04 (p >0.05)
1000	9.70 (σ=5.02)	9.86 (σ=6.02)	**9.93 (σ=5.75)**	H = 0.01 (p >0.05)
10000	**10.11 (σ=6.49)**	**10.11 (σ=6.49)**	**10.11 (σ=6.49)**	H = 0.00 (p >0.05)

Table 6.1: Average number of groups viewed using different ordering algorithms

The Kruskal-Wallis H test has been used to test the significance of these results, and find that for any number of comments up to 30 the results are significant (p <0.05).

As news consumers tend not to read every part of online news content (Holmqvist et al., 2003), this suggests that using this algorithm can increase the number of groups encountered by a large number of news readers.

6.3 Selection of Stories

During the study, participants read the story and associated comments for three news articles. The stories were selected from those collected in Chapter 5 and were selected to be representative of the collection (coming from the most common categories, and having an average number of contributing groups).

Of the 45 stories collected, thirteen concern sport, seven world-news, five showbusiness/celebrity news, and four domestic politics. Other stories concern travel (three stories), motoring (two stories), and technology (two stories). These categorisations were given by the journalists when publishing the stories.

Figure 6.3 shows a histogram of the number of groups generated by the Louvain algorithm against each of the stories (after removing the major outlier "Thai military takes over in coup – again" which had 42 groups). The histogram shows that in this sample the mode number of groups was 6 and mean 7.

Selecting only stories which have the mode or mean number of groups (6 or 7) results in 15 eligible stories. Limiting this set to only the most common categories (sport, world news, and showbusiness) leaves six stories as shown in Table 6.2. From these, this study uses one from the showbusiness category ("Sophia Loren was told she was not photogenic"), one from world news ("World's most expensive divorce set to halve oligarch's fortune"), and one from sport ("Hamilton hits out at Rosberg as rivalry erupts").

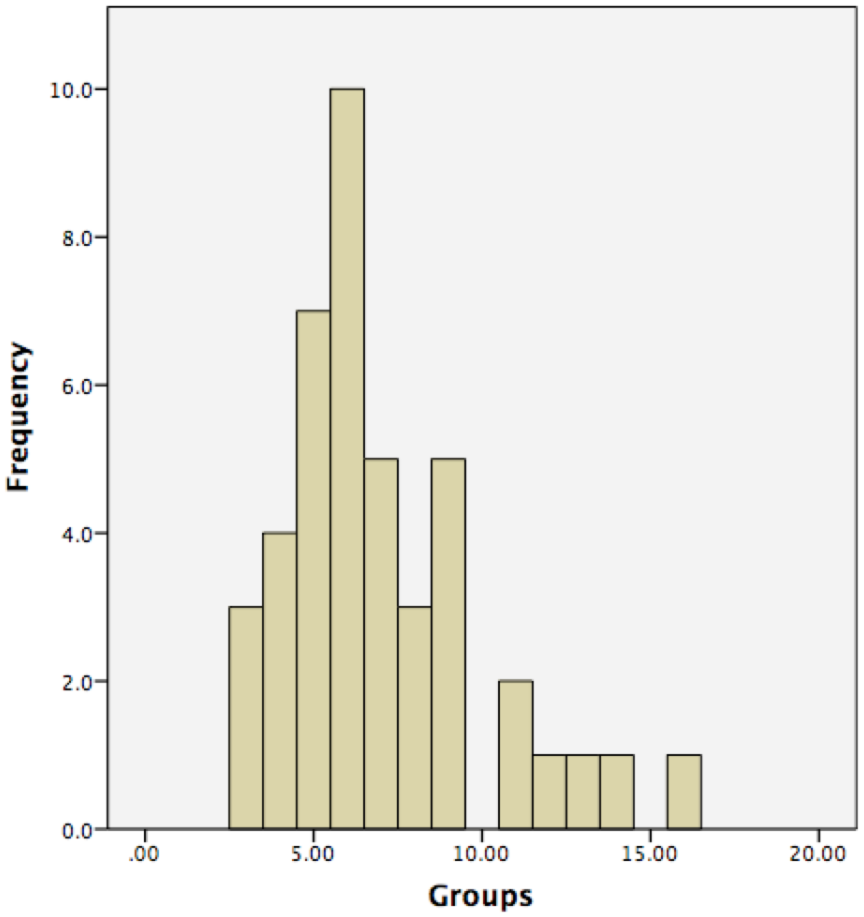


Figure 6.3: Histogram of groups per story (m = 7, σ = 2.96, N = 43)

Throughout the rest of this chapter, the showbusiness story will be referred to as Story 1, the world news story as Story 2, and the sports story as Story 3.

6.4 Experiment Plan

Participants were randomly allocated to one of the four conditions shown in Figure 6.4. Each participant was shown the three news stories identified previously. For participants in Condition 1, the comments were presented using Design A, with the ability to order by chronology or popularity. Participants in Condition 2 were presented with the same interface, however their options for ordering were removed and instead comments were ordered by group to ensure that they read content from the maximum number of groups. Participants in Condition 4 were presented with Design B, which sorts contributions to maximise the number of groups read and allows readers to focus on a single group by selecting it in the summary. Participants in Condition 3 saw the same interface, however instead of using the Louvain Community Detection algorithm, groups were generated randomly. This method of allocating participants allows for distinguishing which changes in perception and behaviour are due to interface changes, algorithm changes, or both.

Source	**Title**	**Category**	**Comments**	**Groups**
CNN	Mounties to check Pamela Anderson's sex abuse allegations	Showbusiness	196	6
Telegraph	Sophia Loren was told she was not photogenic	Showbusiness	129	6
CNN	World's most expensive divorce set to halve oligarch's fortune	World News	135	7
Telegraph	Champions League final is Bale's World Cup	Sport	52	6
Telegraph	Hamilton hits out at Rosberg as rivalry erupts	Sport	375	6
Telegraph	Fabregas wants Premier League return	Sport	198	7

Table 6.2: Stories that concern the most common topics and have either 6 or 7 groups extracted by the Louvain Community Detection algorithm

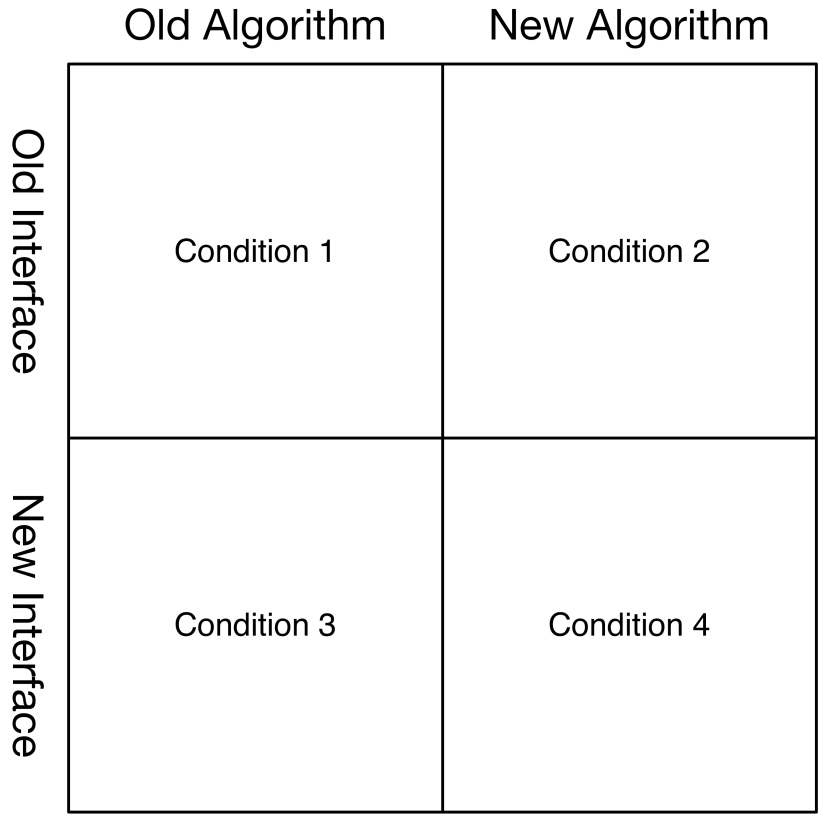


Figure 6.4: Four experimental conditions

For each story, the participants were given four minutes to read the story and discussion. This limit was imposed so that they are not able to read every comment and therefore better simulate typical online news reading behaviours. The limit was initially set to ten minutes but was changed during experimental validation (discussed in Section 6.5). After the four minutes expired, or if they choose to continue early, the participants were asked first to write a bullet points summary of the facts and viewpoints present in the story and comments, and then be prompted to rate the system against Meyer's 5 credibility factors as identified in Chapter 4. Finally, after reading and summarising three stories, they were asked to complete the System Usability Scale. This does not relate directly to the research questions, but will give insight into the usable of such a system and may help explain the results. The System Usability Scale requires scoring ten items on a scale of agreement ranging from 0 to 4, with 0 representing strong disagreement and 4 representing strong agreement (Brooke, 1996). The ten items to be scored are:

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

6.5 Experimental Validation

To ensure that the experiment interface is understandable and that there are no misleading instructions or technical difficulties, a concurrent think-aloud protocol was used whereby two participants were asked to complete the study while verbalising their actions and thoughts (see Lewis and Rieman 1993 for an introduction to the think-aloud protocol). Notes were taken of areas where problems arose, which can be seen in Appendix E.

The first of these sessions revealed that the 10 minute timeline was long enough to read the entire story and all of the comments. This is not how people typically interact with online news content (Holmqvist et al., 2003). Instead people will often scan until finding specific points of interest. To better match this behaviour, the study was modified so that the participants were given only four minutes to read the story and the comments. This shortened time limit is to encourage participants to behave a more natural manner, skimming for the major points of the story, rather than attempting to pick up every aspect.

The participant also spent a large amount of time formatting their summaries, sometimes deleting and rewording sections multiple times. This made the session take longer than it otherwise would have and meant that potentially useful statements were removed during the editing. To alleviate this, the wording of the question was changed so as to ask the participants to "use bullet points" rather than simply to "summarise". When Kriplean et al. (2012) used a wiki to allow participants to summarise comments, they found that participants felt they must commit to summarising everything, and that using a bullet list instead allowed them to highlight the key points which were most salient. This change allowed the gathering of key points without requiring participants to spend a long time writing and formatting summaries.

The participant in the first session also spent the majority of their time, and the majority of their summary, focusing on the article while neglecting the comments. Before the second session the interface was modified so that the article was removed entirely and the participant was presented only with the discussion.

The second session suggested that the changes did help focus the participant on the comments, stopping them from being able to read every comment and encouraging them to skim the article and comments.

The combination of bullet-point summaries and reduced reading time lowered the time of the experiment from over one hour to around 30 minutes. Removing the story did encourage the participant to concentrate on the comments but resulted in a lack of context which caused confusion. To avoid the issue of participants focusing too heavily on the story at the expense of the comments, while still including the story for context, the summary question was separated into two separate questions: one asking for a summary of the points in the story, and another asking for a summary of the points in the comments.

Finally, a small-scale pilot study was held under real study conditions with five participants in order to provide final validation that the study is feasible and will produce results. Five participants is too few to draw any conclusions from the results but this confirmed that participants could complete the study in 30 to 40 minutes, and that they provided meaningful summaries in their responses.

6.6 Results

Data collection was carried out between September and December 2015, involving 63 participants. The demographic makeup was similar to that of the study in Chapter 5, generally young (74.60% younger than 35) with a majority (85.94%) university educated. This study had a higher proportion of students (46.03%) and a more even gender-divide (49.21% male, 50.79% female). For full demographic information see Appendix F. As there are a number of analyses to consider, the results will be presented in this section, before being discussed together later in the chapter.

6.6.1 Usability

In this study, participants using the new interface gave a lower rating on the System Usability Scale than those using the existing interface. Participants using the existing interface gave an average rating of 77.90 (σ = 9.67), whereas participants using the new interface gave an average rating of 70.55 (σ = 13.53). This difference can be explained by the familiarity participants already have with the traditional interface. However, 68 is considered to be the "average result" on the System Usability Scale, so the participants in this study rated both the traditional interface and the new interface as being above-average in terms of usability. This indicates that participants understood how to interact with both interfaces and so differences in results between systems can not be easily explained by differing levels of usability.

6.6.2 Behaviour

The time was recorded when participants started reading the article, started reading the comments, and finished reading. This provides an indication of the length of time spent reading articles and reading comments. Participants on average spent 218.38 seconds (σ = 41.58) reading each story, with 116.17 seconds (σ = 66.35) on the article and 101.75 (σ = 62.34) on the comments. There is no significant difference between those using the old interface and those using the new interface in time spent reading the article (Mann-Whitney U = 4074.5, p >0.05), comments (U = 3967.0, p >0.05), or overall (U = 4138.0, p >0.05).

In condition 3 and 4, participants were first presented with the "all comments" group. When using random groups, participants on average changed the displayed group 1.22 (σ = 2.2) times per story, whereas those using generated groups changed an average of 1.47 (σ = 2.18) times. This difference is not significant (U = 981.5, p >0.05).

	Four Conditions	**Interface**	**Algorithm**
Fair	H = 3.477 (p >0.05)	U = 4828.500 (p >0.05)	U = 4140.000 (p >0.05)
Unbiased	H = 4.437 (p >0.05)	U = 4943.000 (p >0.05)	U = 4183.500 (p >0.05)
Whole Story	H = 3.469 (p >0.05)	U = 4435.000 (p >0.05)	U = 3623.500 (p >0.05)
Accurate	H = 3.162 (p >0.05)	U = 4795.500 (p >0.05)	U = 3896.000 (p >0.05)
Trust	H = 4.364 (p >0.05)	U = 4666.000 (p >0.05)	U = 3824.500 (p >0.05)

Table 6.3: Kuskal-Wallis H and Mann-Whitney U when comparing credibility ratings across conditions

6.6.3 Credibility

Meyer's credibility index (Meyer, 1988) is again used to measure credibility, as it was in Chapter 4. As the factors being compared are ordinal, the Kruskal-Wallis H test is used to to identify if there is a difference between the four conditions, and Mann-Whitney U to identify if there is a difference between the old interface and new interface or between the old algorithm and new algorithm. This will show if the changed presentation brings social news closer to traditional news in terms of perception, and will ensure that any change revealed is due to the sentiment groups and not solely due to the changed design.

The full results can be seen in Appendix G, and the results are summarised in Table 6.3. Despite my expectations the results showed no significant difference on any measure when changing interface, or when changing algorithm. This result is discussed in Section 6.7.

6.6.4 Reasons given for ratings

As giving reasons for ratings was optional, only 93 submissions provided reasons (out of 189 in total). Those provided had an average of 37.48 (σ = 28.49) words and 211.28 (σ = 162.29) characters.

The reasons were organised into an evolving set of categories, and I revisited and refined these categories until all reasons had been allocated. These categories are listed in Table 6.4 and all reasons are listed in Table 6.5. The breakdown across stories can be seen in Appendix H. Features of the article were mentioned in 76 of the 93 submissions, whereas features of the comments were mentioned in only 27.

The reasons given fit into three broad categories: reasons related to general attitudes towards news, related to the content of the article, or related to the content of the comments. There were also six submissions that indicated a lack of previous knowledge about the subject influencing their ratings. No submission gave previous impressions or expectations of comments as a reason for their ratings.

Interface **Algorithm**	**Old** **Old**	**Old** **New**	**New** **Old**	**New** **New**	**Old**	**New**	**Old**	**New**
Article content	**28**	18	13	17	**46**	30	**41**	35
General attitude towards news	**9**	4	2	1	**13**	3	**11**	5
Comment content	8	4	3	**12**	12	**15**	11	**16**

Table 6.4: Credibility Reason themes. Grouped by old/new interface, old/new algorithm, and combinations of interface and algorithm.

6.6.5 Interviews

To further investigate the reasons for the results, detailed interviews were undertaken with eight participants, two from each of the four conditions. The interviews were semi-structured around eight questions:

1. What elements did you consider when evaluating the credibility of the story?
2. What elements do you feel enhanced credibility?
3. What elements do you feel detract from its credibility?
4. What do you think is missing that would make this feel more credible?
5. How did the comments change your opinion on the original article?
6. Did you feel there was a meaningful order to the comments?
7. Do you feel the comments fairly represented the range of views on this topic?
8. How did you interpret the five credibility criteria?

These questions were produced after analysis of the behaviour and credibility data, and the reasons given. The lack of significant changes between conditions lead me to consider two potential causes: that highlighting disagreements in the comments (by promoting different viewpoints) results in some participants viewing the overall story as less credible and others viewing it as more credible, or that participants generally did not consider the comments to be an important part of the story. The eight questions were chosen to provide insight into which of these is most likely. The purpose of the first four questions is to understand if there were particular features of the interface or the articles that affected the ratings. The purpose of the next three is to understand how much attention readers actually paid to the comments. With the final question, the participants' understanding of the credibility criteria was examined and compared to their answers about features of the story and comments, and to the ratings they provided.

Interface	**Old**	**Old**	**New**	**New**	**Old**	**New**		
Algorithm	**Old**	**New**	**Old**	**New**			**Old**	**New**
Journalist/Editor is biased	**5**	1	2	4	**6**	**6**	**7**	5
Comments disagree with article	**1**	0	**1**	0	**1**	**1**	**2**	0
Mentioned comments	8	4	3	**12**	12	**15**	11	**16**
Lack of knowledge/information about subject	**6**	0	0	0	**6**	0	**6**	0
General lack of trust in news	**4**	1	2	1	**5**	3	**6**	2
Sensationalised	1	3	**4**	0	**4**	**4**	**5**	3
Missing information in article	**11**	4	5	8	**15**	13	**16**	12
Expectation for news to be true	**4**	3	0	0	**7**	0	**4**	3
Dispute in comments	1	0	0	**2**	1	**2**	1	**2**
Story is simple retelling of facts	3	**4**	1	2	**7**	3	4	**6**
Mentioned groups	0	0	2	**5**	0	**7**	2	**5**
News outlets are biased	**4**	1	0	1	**5**	1	**4**	2
Mentioned headline	**2**	**2**	1	1	**4**	2	**3**	**3**
Comments are biased	1	1	0	**3**	2	**3**	1	**4**
Things in comments not mentioned in article	1	1	0	**3**	2	**3**	1	**4**
Lack of perspectives	**5**	1	3	4	6	**7**	**8**	5
Comments not accurate	1	1	0	**2**	**2**	**2**	1	**3**
Representative of perspectives	1	0	0	**3**	1	**3**	1	**3**
Mentioned article	**25**	17	13	16	**42**	29	**38**	33
Number of reasons	9.524 (σ=3.507)	8.000 (σ=2.649)	6.255 (σ=2.743)	8.089 (σ=3.817)	8.688 (σ=3.159)	7.115 (σ=3.415)	7.731 (σ=3.511)	8.042 (σ=3.250)

Table 6.5: All Credibility Reasons. Grouped by old/new interface, old/new algorithm, and combinations of interface and algorithm.

The key points of each interview are listed below. Extended summaries and full transcripts are available in appendix J.

6.6.5.1 Condition 1: Old Interface Old Algorithm

	Participant 1			Participant 2		
Ratings						
	Story 1	**Story 2**	**Story 3**	**Story 1**	**Story 2**	**Story 3**
Fair	4	1	3	3	0	2
Unbiased	4	1	3	3	0	1
Whole Story	4	2	2	2	0	1
Accurate	3	3	2	2	1	1
Trust	4	2	2	2	1	2
Key Points						
	• Looked for quotations and references. • Too many quotations from a single source detracts from credibility. • Would liked to have known the news source. • Ran out of time reading comments. • Comments were unrelated to the story. • Comments were quite biased. • Disregarded comments. • Did not sense any order to comments. • Some comments referenced third party sources, establishing credibility.			• The journalists' viewpoints influenced articles. • News outlet is important. Would typically consult multiple outlets. • Did not sense any order to comments. • The comments represented multiple viewpoints, which enhanced credibility. • Primarily considered the article - though comments influenced in a small way.		

Table 6.6: Condition 1 interview credibility ratings and comments

6.6.5.2 Condition 2: Old Interface New Algorithm

	Participant 3			**Participant 4**		
Ratings						
	Story 1	**Story 2**	**Story 3**	**Story 1**	**Story 2**	**Story 3**
Fair	2	3	2	1	3	1
Unbiased	2	3	1	2	3	1
Whole Story	1	2	1	2	3	0
Accurate	1	3	2	1	2	1
Trust	2	3	2	2	3	1
Key Points						
	• Basic facts of stories were believable, easily verifiable. • Poor writing style detracted from credibility. • Comments did not contribute much due to fact-based stories. • For Formula One story, comments restated facts in a more accessible way. • No meaningful order to comments. • Comments represented a wide range of viewpoints, though overrepresented some. • Concentrated on the article when providing ratings, paying little attention to comments.			• Contradictions between comments and article harm credibility. • More credible when comments support article. • Stories which leave open questions lack credibility. • No meaningful order to comments. • In some cases viewpoints were fairly represented, in others some were excluded. • After reading top comments, other comments repeated what had been said. • Compared comments to article when deciding credibility.		

Table 6.7: Condition 2 interview credibility ratings and comments

6.6.5.3 Condition 3: New Interface Old Algorithm

	Participant 5			Participant 6		
Ratings						
	Story 1	**Story 2**	**Story 3**	**Story 1**	**Story 2**	**Story 3**
Fair	3	3	1	2	2	2
Unbiased	3	2	1	2	1	2
Whole Story	3	2	2	3	1	2
Accurate	3	3	1	2	2	2
Trust	1	0	0	2	2	2
Key Points						
	• These are entertainment stories, so will have some element of bias. • Looked for specific facts and details to provide confidence of credibility. • Too many opinions result in a less credible story. • Many comments were not relevant to the story or based on facts from the story. • Had difficulty assessing the credibility of the Formula One story due to lack of knowledge about Formula One. • Ran out of time reading comments. • Concentrated on the article rather than comments. • Did not notice a difference when moving between tabs. • Knowing the identity of the journalist or news organisation would have made credibility assessments easier.			• Difficult to judge credibility of divorce story due to lack of detail. • Credible stories shouldn't exclude any "really clear" facts. • Newspapers editorialise to make stories more exciting. • Comments tended to take the same angles as the articles, though in some cases they added information. • Valued disagreements in the comments to provide additional points of view. • Understood groups, but primarily used "View All" section. • Did not perceive an order to the comments. • Comments covered a variety of viewpoints, though were more biased than the article. • Looked to the source of information (journalist, agency, source of quotations) etc. to provide confidence of credibility.		

Table 6.8: Condition 3 interview credibility ratings and comments

6.6.5.4 Condition 4: New Interface New Algorithm

	Participant 7			Participant 8		
Ratings						
	Story 1	**Story 2**	**Story 3**	**Story 1**	**Story 2**	**Story 3**
Fair	1	1	1	4	4	4
Unbiased	1	1	1	3	4	4
Whole Story	2	0	0	3	4	4
Accurate	2	1	1	3	4	4
Trust	1	1	1	4	4	4
Key Points						
	• Focused on the source of facts in the article. • General doubt about the credibility of news stories. • Direct quotations enhance credibility. • Multiple angles enhances credibility. • The comments primarily focus on areas the article did. • Perceived no meaningful order to the comments. • The comments fairly represented the range of views. • Rated purely on the article and didn't take the comments into much consideration.			• Placed a lot of trust in news outlets. • Journalists may have their own biases. • Did not feel the comments influenced them much. • Strong difference between "really biased" comments and journalistic news. • Perceived no meaningful order to the comments. • Felt there was too many groups, overwhelming with information. • The comments fairly represented the viewpoints. • Based credibility assessments on article similarity to stories they'd typically read in outlets they trust.		

Table 6.9: Condition 4 interview credibility ratings and comments

6.6.6 Summaries Written by Participants

Participants provided article summaries that were on average 45.61 words ($\sigma = 31.94$) and 271.33 characters ($\sigma = 182.78$) long, and comment summaries that were on average 35.92 words ($\sigma = 28.43$) and 211.25 characters ($\sigma = 163.35$) long.

All article and comment summaries were read and commonly mentioned topics were identified. These were refined to merge similar topics into appropriate categories and this was repeated until all major topics were identified and the appropriate summaries

were tagged. This resulted in 20 topics for Story 1, 20 for Story 2, and 13 for Story 3. The distribution of the topics to summaries can be seen in Appendix I. On average, article summaries mentioned 3.78 ($\sigma = 1.69$) topics and comment summaries mentioned 2.88 ($\sigma = 1.93$). There are correlations between the time spent reading an article and the number of article topics contributed (Pearson's correlation coefficient $r = 0.3$, p <0.05) and between the time spent reading comments and the number of comment topics contributed ($r = 0.44$, p <0.001).

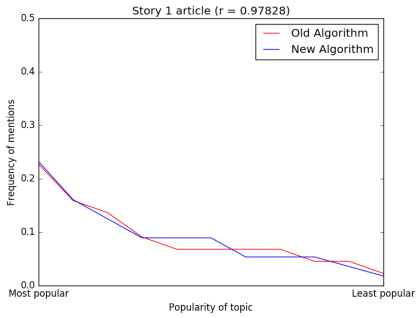
As this study involved independent measurements, there is a risk that natural differences between participants (for example differences in the verbosity of their summaries) will result in differences in summaries that are not reflective of changes due to the system. However, there was no significant difference between those participants using the new algorithm and the old algorithm in article summary length (U = 4161.0, p >0.05), comment summary length (U = 4016.5, p >0.05), article topics mentioned (U = 4482.5, p >0.05), or comment topics mentioned (U = 5002.5, p >0.05).

The topics identified in the summaries indicate the topics that participants believed were most important, but the lack of mention of a topic can not be taken to indicate that the participant did not consider or encounter that topic. For this reason, this section will compare the relative frequency of topics across conditions rather than their absolute number of mentions. In Figure 6.5, all topics have been ordered by usage and plotted on the X-axis. This means that points towards the left side of the graphs represent the topics that were most commonly mentioned by participants in that condition, but does not mean that these are the same topics across conditions. To identify any difference in relative popularity between conditions, Pearson's correlation coefficient (r) is used.

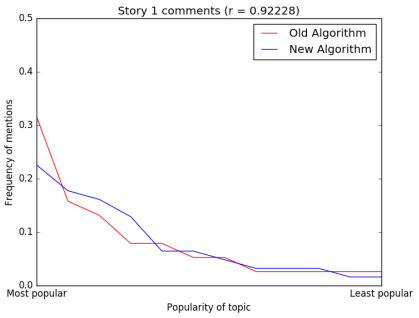
These figures show a very high level of correlation between the relative popularity of topics mentioned by participants using the new algorithm and those mentioned by participants using the old algorithm for both articles and comments. These results are significant (p <0.01). These results do not suggest that the change in algorithm has resulted in any difference in the relative popularity of the most popular topics in each condition.

As well as investigating the relative popularity of topics, I am also interested in the different levels of popularity of specific topics. For this the topics are again ordered by popularity within each condition, and Levenshtein distance [3] is used to compare the similarity of orders between conditions. If the change in algorithm results in a change in the frequency of mention of specific topics, then the Levenschtein distance between topics mentioned in comment summaries in different conditions would be higher than the distance between topics mentioned in article summaries.

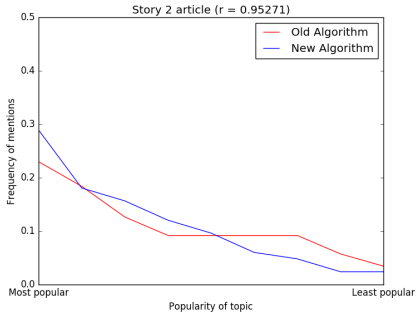
[3]Levenschtein Distance measures the level of similarity between two sequences (the minimum number of edits needed to turn one into the other, with the allowed operations being inserting, removing, or replacing an element of the sequence). A lower number indicates a more similar sequence.



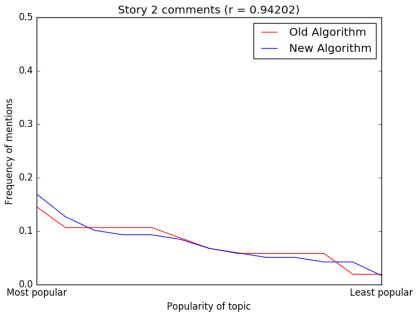
(a) Story 1 Article



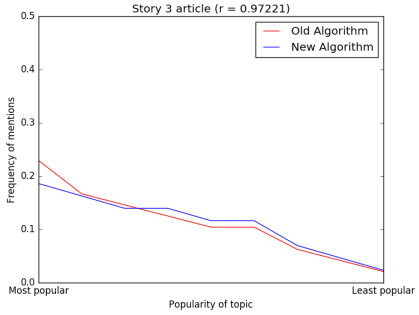
(b) Story 1 Comments



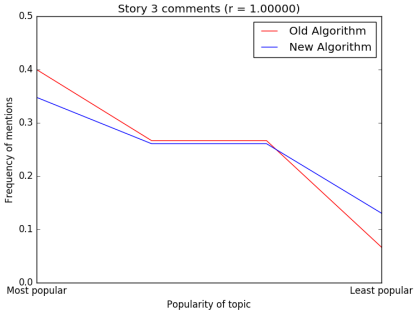
(c) Story 2 Article



(d) Story 2 Comments



(e) Story 3 Article



(f) Story 3 Comments

Figure 6.5: Topic frequency across story and comment summaries, with Pearson's correlation coefficient

As comments and articles have different numbers of topics (and a larger number of topics allows for a larger Levenschtein distance) the Levenschtein values are normalised by dividing by the number of topics. The results are shown in Table 6.10.

Table 6.10: Distance in ranking of topics between old algorithm and new algorithm

Story	**Article Distance**	**Normalized**	**Comment Distance**	**Normalized**
1	7	0.54	8	0.53
2	4	0.36	10	0.56
3	5	0.45	6	0.55

This shows a higher distance between article and comment topics when using the new algorithm in two of the three stories, and a very similar level of difference in the other. With only three stories no conclusions can be drawn about significance or larger trends.

6.7 Discussion

The majority of measures used in this study showed no significant difference between using the old or new algorithm, or between using the old or new interface. In this section I will review the results and explore possible explanations for the lack of significant changes.

During this study the participants in condition 3 and 4 (using the new interface) changed the selected tabs a relatively low amount (1.22 times per story for random groups and 1.47 times when using generated groups). This suggests that most participants were primarily interacting with the default selected "show all" tab. This means that participants in condition 3 primarily viewed a randomly ordered set of comments, and those in condition 4 primarily viewed the same set of comments as those in condition 2, though with the additional context provided by colour-coding the commenter groups. Though the participants did not typically view a large amount of groups, they did spend significant amounts of time reading comments. Almost half of participants' time (101.75 seconds per story) was spent reading comments.

Additionally, the interviewed participants were able to explain the difference between the five credibility factors, and could explain the elements that they used when judging those factors. They typically viewed fairness and being unbiased as being very similar, with both concerned with promoting multiple viewpoints and perspectives. To assess if they have been told the whole story they typically attempted to identify unanswered questions or unexplored topics, and made a judgement on if the amount of content seemed reasonable for the topic. For accuracy, some interviewees mentioned that they would typically check other outlets to help decide this factor. However in this case they could not do that and they used a number of ways of assessing the accuracy to compensate. This included features which were a combination of how fair and unbiased the article was, statements that can be compared to prior knowledge, and the presence of named sources and quotations. Finally, for trust, several interviewees again mentioned they would typically compare multiple sources to assess this. In the absence of information about the source, two participants claimed they had great difficulty making the judgement, two compared the styles of writing to news media they know, one looked to the comments for reference to other news sources that can back up the story, two based it on a combination of other factors, and one looked for disagreements between the article and the comments.

The System Usability Scale ratings alongside this behavioural data confirm that participants understood and engaged with the system. The interviews confirm that participants had an understanding of the metrics they were assessing the articles by.

Chapter 5 verified that by using community detection algorithms the system is able to group together content in a similar way to a human, and the evaluation earlier in this chapter confirmed that presenting the information using this interface will result in a greater number of groups being encountered. This chapter has confirmed that participants understood and engaged with the system presented and had an intuitive understanding of the credibility metrics. However, despite this there was no significant difference for any of the credibility factors when comparing between conditions, interfaces, or algorithms.

Upon initial analysis, two potential reasons for this emerge. Either any positive impact of the presented comments were matched by equal negative impacts, or participants did not consider the comments to be part of the article when they were assessing credibility.

In support of the first possibility, the reasons given alongside ratings do include positive and negative points made about both the story and the comments. Some participants had a basic expectation for news stories to be true (e.g. "I would have expected some truth to be represented in a news article" and "I would expect the news to give me a fully rounded view of the story") whereas others expected the opposite (e.g. "News stories only share what we would like to hear", "In general I do not believe everything I read."), and others commented on specific aspects of the article that influenced them positively (e.g. "It appears to give an objective assessment of the facts, and as such can be trusted", "Fairly short, objective article. Mostly reporting facts that could be verified fairly easily") or negatively (e.g. "I feel that this article is extremely biased", "seems very biased from the beginning of the article").

There were also reasons given regarding both positive and negative aspects of the comments (e.g. "because there was a clear dispute between the yellow group and blue group there was a sense of balance to the reporting", "comments presented were clear and seemed representative of the different views", "the comments may be uninformed.", "the comments [...] are not supported by evidence"), though none of the reasons given related to overall impressions of commenting systems.

This mixed opinion of comments was also present in the interviews. Participant 4 said that they "compare[d] the article with the comments [to] get the whole story.". Participant 1 claimed the comments helped them trust the story as they "[say] something about that news is truth and is not fake", whereas participant 5 believed that the comments they read were full of opinion not based in fact, and even that the comments "made [the participant] forget what the actual story was about".

However the second possibility, that the comments were not considered to be a major contributor to the credibility of the story, is well supported.

First, features of the article were mentioned in 76 of the 93 submissions, with only 27 mentioning features of the comments. Second, when interviewed, participants primarily referenced features of the article rather than comments when questioned about credibility. For example participant 1 saying that they would have been more convinced of the credibility if there was "more references from other news agencies", and participant 2 referring to their preference in news websites when assessing the stories ("If you visit a credible website like BBC you would assume that the information would be true, but if you visit the Daily Mail youd probably assume that they're not telling the whole story").

Additionally, several of the participants interviewed acknowledge that they were primarily focusing on the article when deciding credibility. For example participant 3 said "I don't know that the comments add a lot.", participant 1 said they disregarded many comments as they were "not related to the actual news", participant 8 said that comments "wouldn't influence [them] because those will be really biased" and when asked if they were taking the comments into account, participant 2 said they were focusing "mostly [on] the article itself".

This tendency to neglect the comments when providing ratings is also clear when comparing the ratings given by interviewees to the reasons given during the interview. Several participants gave conflicting answers, indicating that they felt a particular aspect was important but not reflecting that in their ratings when that aspect was present in the comments. For example participant 7 believed that equal representation of viewpoints is important for both fairness and bias and stated that the comments did fairly represent the range of viewpoints, but still gave all three stories low ratings for fairness and bias. When this was mentioned during the interview they acknowledged that when rating they did not take the comments into as much consideration as the article ("I think maybe when I was rating it I'm rating it purely on the story and not taking the comments into as much consideration.").

In Faridani et al.'s evaluation of their Opinion Spaces platform, the authors found that the feedback suggested that users "want to better understand the arrangement". The interviews found that despite the introduction to the study explaining the groups, and at least one participant claiming that they "understand the concept and can tell that they share the same point of view", none of the eight interviewees sensed any meaningful order to the comments (other than chronology). This may have also contributed to the lack of change in credibility rating.

The topics that people mentioned during summarising the articles and comments provide no evidence of a change in the relative popularity of topics contributed by participants in different conditions, however they do suggest that there may be qualitative differences

to the topics mentioned in different conditions. As this study is too small to determine the scale of this change with any certainty, this should be a question for future study.

Earlier in this chapter a weakness in the methodology was identified: participants are being presented with news comments rather than complete news stories. Using this methodology provided no evidence that the change in algorithm results in a change to perceived credibility or to the relative popularity of topics mentioned in summaries, though there was some evidence of a change in which topics are mentioned in summaries.

The reasons given for the ratings, and the additional information given during the interviews, give no reason to believe that using data from an existing social news system rather than the discussion system used in this study would have produced different results. This is because social news systems often present content in the context of a third-party story, and so it is likely that even if this experiment was performed using existing social news systems' data, the participants would still focus primarily on the article rather than the comments.

Applying the algorithm to a system where the actual story itself is contributed by multiple named authors and so the content can be organised to optimise representation may yield different results. However, no systems like this currently exist so any experiments would need to be carefully planned to ensure that results do not just reflect the novelty of the overall system. An experimental plan and prototype system for such a study is presented in Appendix K as future work.

6.8 Conclusions

This chapter first documented the use of an iterative participatory process to design two interfaces for news discussion. Each of these designs began with inspiration from existing news systems and were then refined through participatory sessions with stakeholders.

The focus of one design is to highlight the structured information gained by the use of the community detection algorithm in the expectation that this will change reading behaviour. The final version of this design presents the groups in a very prominent location at the top of the discussion, colour-codes contributions according to groups, and provides links to discover more information and contributions from any given group. The second design accurately represents the state of online discussion interfaces in use today. This includes presenting votes in aggregate (as opposed to individual up and down votes), and ordering by chronology or popularity.

A preliminary evaluation of the new design was performed, confirming that ordering the contributions according to group allows for presenting a larger number of groups to readers in the same number of comments. I then presented the plan and execution of a study of 63 participants to investigate the effects of grouping algorithms on the way

that people consume news content. It was established that participants did engage with the system, spending on average almost half of their time reading the comments, though those presented with the new interface changed the displayed group a relatively low number of times. This section will evaluate these results in respect to the hypotheses.

H1. Adding structure using the technique detailed in Chapter 5 will result in an increase in credibility

In this study there was no significant difference in the perceived credibility of the stories when changing interface or when changing algorithm.

The most likely reason for lack of impact appears to be that the participants focused their attention on the article rather than on the comments, so the novel features of the comments failed to make an impact. This is shown in the reasons participants gave alongside their ratings, which primarily focused on the original article itself rather than the comments. There were also a number of participants who gave previous experiences of news as reasons for ratings but none who mentioned previous experiences with comments.

Through semi-structured interviews with eight participants, most participants indicated that they were primarily evaluating the credibility of the article and not the comments. The reasons for this varied, with some feeling that the comments had little to offer in terms of the story, and others acknowledging that the comments made a contribution to the story but still failing to account for that in the ratings.

H2. Adding this structure will result in an increase in the number of points of view encountered during reading

To determine if the change in algorithm resulted in a difference in the perspectives considered, the topics most commonly mentioned in the summaries were identified and the relative popularity of topics was compared between article summaries and comment summaries.

This showed no change to the relative popularity of topics within conditions, though it did suggest that it may result in a change in which topics are mentioned. As only three stories were analysed in this study, larger conclusions about this cannot be drawn.

The study in this chapter found that after being instructed to treat the article and the comments as one complete story, and despite participants understanding the criteria for credibility and in some cases deciding that the comments met this criteria, participants still gave ratings according to how they viewed the original article or the journalist. It also provided limited evidence that the change in algorithm may result in readers engaging with different topics, though this needs to be further investigated in future work.

These results suggest that it will be very difficult for open news platforms to ever gain the credibility of traditional news outlets as readers already have a strong definition of news that excludes reader contributions, but does suggest that the editors of traditional news outlets can experiment with forms of citizen engagement as poor quality reader contributions do not appear to negatively affect reader views of the overall story.

Chapter 7

Conclusions

This thesis presented work which aimed to investigate the openness of media outlets to citizen contribution, the relationship between openness and credibility, and methods of increasing the credibility of non-traditional media outlets. This was approached by automating some of the functions of a news editor in order to present a wider range of viewpoints to news consumers.

Chapter 2 established the important roles the news media play in society, and reviewed some areas where they are expected to contribute. I also identified areas where they are failing in these roles and the negative effects this can have on the proper functioning of democracy. A lack of perspectives in much media output was identified as one possible cause for these failures, with the emergence of online news platforms noted as a possible solution, allowing for many new voices to contribute to the news.

Chapter 3 proposed a new framework for analysing online news systems in terms of their openness throughout the news process and used it to show that traditional news outlets maintain control of most of the news process, and that even citizen news systems only open some, whereas the systems that do open most of the process are not typically considered to be news sites at all (e.g. YouTube or Reddit).

Chapter 4 presented a study involving 79 participants rating news systems in terms of credibility. The ratings provided were then compared to the news systems' position on the spectrum of openness. This found that openness is inversely correlated with the perceived credibility of a news system, suggesting that though open news systems are becoming more commonly used, some features of these systems result in them lacking the credibility of traditional news outlets. This may reduce the effectiveness of these systems in increasing the range of voices present in the media and suggests that research should investigate ways of creating credible, open, news systems.

Chapter 5 identified a method of using community detection algorithms to group together contributors in news discussions who share similar sentiments, with a goal of

using this to automatically add structure to unstructured news systems, and partially perform the tasks of an editor, with an eventual goal of increasing the credibility of these systems. Through an experiment, it was shown that significantly more people voted that two comments share "similar sentiment" when those two comments were from authors that the system grouped together. The study in this chapter also showed that when the grouping is treated as a vote, the system agrees with people almost as often as an average person does.

Finally, Chapter 6 documented the design of an interface to present comments associated with news stories so as to maximise the displayed groups. This interface was presented to 63 participants and it was found that the novel presentation did not change their perceived credibility of the articles, though there was some limited evidence that it may lead to them encountering different perspectives. Through analysis of the reasons given for their ratings, and through semi-structured interviews with a number of participants, it was found that the likely reason for this is that the participants saw the story as being the article written by the journalists, with the comments being something separate and not integral to the story.

7.1 Contributions

The contributions made in this thesis are:

1. A new framework for analysis of online news systems (Chapter 3)
2. A spectrum of openness of 32 online news systems (Chapter 3)
3. Four case studies showing the flow of news between systems (Chapter 3)
4. A credibility study covering 27 online news systems (Chapter 4)
5. An application of the Louvain Community Detection algorithm to online news discussion systems, and its successful evaluation (Chapter 5)
6. A design for presenting structured news discussions (Chapter 6)
7. An analysis showing that additional structure does not result in a change in credibility, but may result in a change in perspectives encountered (Chapter 6)

7.2 Research Questions

This section will review the contributions made in this thesis in the context of the research questions presented in the introduction.

7.2.1 How might we measure openness in terms of citizen participation in the news process?

In Chapter 3, this research question was broken down into two sub-questions:

1. How much power is being given to citizens through online news systems?
2. What sort of interactions exist between those systems that are more open and those that are less?

The first of these was answered using a modified version of the methodology used by Domingo et al. (2008) to analyse online news websites. A spectrum of openness was produced which showed that traditional news organisations do not provide much power to citizens, instead retaining overall control of the news process. This is true even in systems that are explicitly aimed at gathering citizen contribution, where instead of handing control to readers they typically attempt to recreate activity that is already happening on the web within a single system they control. However, the spectrum did show that alternative news systems do exist which do transfer that power, including social networks as well as explicit social news platforms, blogging platforms, and alternative news structures such as Wikinews. Though these alternative systems do allow for citizen participation at every stage of the news process, they lack the authority of traditional news media.

The second question was approached by utilising case studies of four stories that travelled through a number of news systems from the spectrum. The case studies were selected to represent varied types of online news and included one request for contribution from a major news publisher, a campaign instigated by the members of an online forum, a video that became viral and was reported on in traditional media, and a story in a mainstream outlet that was then discussed and responded to on social media and blogs. These case studies revealed a far more complex situation than presented by the spectrum. They showed that stories are not constrained by the spectrum, and instead they move between systems as they progress through the stages of news production. This process includes moving from the more open systems to the more closed (as in the first three case studies), as well as moving from closed systems to open systems (as in the final case study).

7.2.2 Does a relationship exist between the openness of a system and its credibility?

In Chapter 4 this question was approached through a study of 79 people and their perceived credibility of the systems from the spectrum. This involved the participants indentifying systems that they are familiar with and then rating those systems using the credibility criteria proposed by Meyer (1988).

The results of this study showed that each of the credibility factors has a modest but significant negative correlation with openness (Fairness -.285, Unbiased -.307, Whole Story -.272, Accurate -.349, Trustworthy -.315. ps <0.01). This correlation exists whether or not responses are included from participants who have not interacted with the system in question.

Analysing the comments left by participants provided evidence that the reason that two of the systems (The Sun and The Daily Mail) were both closed and lacking credibility is either because of their reputation for entertainment rather than factual news, or because of their strong ideological positions causing strong reactions from the participants. Comments also showed disagreement in whether the more open systems can be rated as a whole and even if they can be considered to be news at all.

7.2.3 Will promoting a wide range of viewpoints (using viewpoint groupings) change the exposure of readers to different viewpoints, and alter the perceived credibility of a system?

Chapter 5 identified an algorithm capable of identifying groups of users who share sentiment in a graph of social behaviour from a news discussion system. The purpose of this is to add enough structure to news discussion so as to be able to fulfil some of the responsibilities of news editors as specified in Chapter 2: stimulating debate amongst readers, ensuring objectivity and balance, and avoiding duplication.

Chapter 5 showed that the sentiment groups generated by the algorithm are meaningful, with comments by authors allocated to the same group making comments rated "similar" 66.91% of the time, and authors allocated to different groups making comments rated "similar" 43.6% of the time. I also found that by using them as an indication of the sentiment of posts by those users, the system can make judgements which agree with people almost as often as an average person (the system agrees with people 62% of the time, compared to 66% for an average person, and 45% expected from random groupings).

Chapter 6 documented the design of an interface for presenting these groups to users and confirmed that when reading up to 30 comments, readers using this interface will be exposed to a greater number of groups than those using traditional interfaces ($p < 0.05$). This chapter also used the interface to investigate the effects on the way that people consume news discussions. This was approached by investigating if adding structure using the technique detailed in Chapter 5 would result in an increase in credibility.

The study found that presenting the comments in this novel way resulted in no significant change to the credibility ratings assigned for any credibility criteria. This was true when changing interface, algorithm, and when changing both. It was determined that the most likely reason for this is that the participants did not view the comments as

being as important as the journalistic article, and though they did acknowledge that the comments fulfilled their criteria for credible news, this was not enough to change their opinions of the stories' overall credibility.

However, there was some limited evidence of change in the viewpoints that people encountered when reading the comments. This was found when comparing which topics were mentioned most often in summaries by participants in different conditions. However, because only three topics were used in this study no firm conclusions can be drawn from this and future work should investigate this further.

7.3 Recommendations

Perhaps the strongest result from this work is that despite participants being instructed to consider the comments as part of the overall story, and despite acknowledging that their criteria for credible news were partially fulfilled by the comments, the participants did not account for the comments in their credibility ratings. This suggests that though overall trust in the media has been falling over time, readers are still able to differentiate strongly between content that is professionally produced and that which is contributed by other readers.

The findings suggest that it will be very difficult to produce citizen news platforms that have comparable credibility to traditional news media. This work showed that adding additional properties to highlight positive features of reader contribution does not make a difference to the credibility of the story, even when the participants acknowledge those positive features.

These findings reinforce the socially-constructed nature of news, and suggests that those who make news should be aware of the special status news consumers attribute to journalists. They suggest that readers will maintain a strong separation between "primary" news content and secondary content produced by non-journalists, even if in principle the content is very similar. Though the changed interface did not result in an increase in credibility, it also did not result in a decrease in credibility even among those participants who acknowledged non-credible comments. This suggests that journalists can afford to experiment with new forms of citizen contribution without risk to their own reputation, as long as that contribution can be clearly separated from the journalists' primary content.

The limited evidence that the change in algorithm resulted in a change of viewpoints encountered should be explored in future work, as this may suggest that it is possible to expose readers to more varied viewpoints albeit without a resulting change in the overall credibility of the news system.

7.4 Future Work

The work in Chapter 6 made progress towards answering the research question "Will promoting a wide range of viewpoints (using viewpoint groupings) change the exposure of readers to different viewpoints, and alter the perceived credibility of a system?". It suggested that using the method detailed did not result in a change in credibility but may result in a change of viewpoints encountered. However, due to issues of time, I did not attempt to investigate how this change would affect the way that people *contribute* to online news.

If contributors were made aware of the algorithm being used to display their content, it is possible that they may participate in different ways: attempting to manipulate the group into which they are placed, producing unpopular content, or interacting with contributions from less popular groups. Although this was out of scope for this thesis, a study to investigate impacts of the changing algorithm on the way that people contribute to online news discussions has been designed and the software implemented. For the full design of this study, see Appendix K.

Additionally, the study in Chapter 6 was performed using discussions that were created by users of an existing online discussion system. News content generated using this novel interface would potentially have different characteristics to content generated using existing discussion systems, and this may result in changing reading behaviour. The experiment detailed in Appendix K would create a suitable dataset for future research in this area.

Ideally, such an intervention could be made with an existing community in a similar way to the work performed by Ludford et al. (2004). The results in this thesis suggest that this would result in readers viewing, and potentially engaging with, an increased number of viewpoints. The system in Appendix K could be modified for such purposes.

Chapter 5 made note of the inability of the participants to agree on the credibility of different news sources. This result points towards the participants holding different conceptions of "dissimilar sentiment" and may be related to the level of dissimilarity between the sentiment of the comments and that of the reader.

Though not investigated in this work (as participants did not indicate their agreement or disagreement with content and as such could not be placed into groups) it is plausible that knowledge of the viewpoints of readers (through their position in sentiment groups) will allow for predicting their perceptions of similarity of content.

This could lead to the production of systems that generate groupings of users according to the preferences of the reader - allowing for more nuanced groups in areas where that nuance is best understood, and more general groupings elsewhere. Previous work has investigated the the existance and impact of filter bubbles in online media (e.g. Pariser

2011) and the amount of differing opinions readers will tolerate (e.g. Munson and Resnick 2010). Further research in this area could lead to the development of systems that avoid the problem of filter bubbles while ensuring that "challenge-averse" readers (see Munson and Resnick 2010) are not dissatisfied with the content.

7.5 Recent Developments

In the time since the creation of the spectrum of openness (see Chapter 3), there have been a number of changes to systems on the spectrum. Of particular interest, a number of the systems on the right side of the spectrum (which were not typically considered to be news systems) have invested more heavily in portraying their platforms as news systems. For example, in April 2016, Twitter moved its iPhone App from the "Social Network" category of Apple's App Store to the "News" category[1], where Reddit was already listed, and in 2014 Facebook launched their "Trending topics" section which shows lists of the most popular topics being discussed on the platform with a short description of why they are "trending". In 2015 they launched "Live Articles" which allows news publishers to host their content directly with Facebook. These moves re-enforce the findings of Chapter 3, that though the more open systems on the spectrum may not typically be considered to be news systems, they are still used for implementing the news production process.

Though Facebook claims that their trending topics are selected algorithmically based on "factors including engagement, timeliness, pages you've liked and your location"[2], in May 2016 it was alleged by former employees that curators were able to manipulate the topics on the trending topics section[3]. The employees alleged that whether a topic will be promoted "depends on who the curator is and what time of day it is", and also that they were encouraged to inject big news stories that weren't naturally surfaced by the trending algorithms.

Facebook have denied these allegations, but their plausibility does highlight a problem that was acknowledged in Chapter 3 but not strongly addressed: that it is impossible to verify the "openness" of these platforms when their infrastructure and algorithms are private. Some work has investigated the issues with trust in algorithmic news (see Diakopoulos and Koliska 2016), and future iterations of the spectrum of openness should account for this.

There has also been a large recent rise in attention paid to "fake news" shared on social media [4]. In November 2016, Buzzfeed News reported that towards the end of

[1] See `http://mashable.com/2016/04/28/twitter-news-app-switch`

[2] According to `https://www.facebook.com/help/737806312958641`

[3] See `http://gizmodo.com/former-facebook-workers-we-routinely-suppressed-conser-1775461006`

[4] e.g. "This is how Facebooks fake-news writers make money" in the Washington Post November 18 2016, "How Fake News Goes Viral: A Case Study" in The New York Times, November 20 2016, etc.

the 2016 US Presidential election, "the top-performing fake election news stories on Facebook generated more engagement than the top stories from major news outlets" [5]. Their analysis focused on content with the most shares, reactions, and comments on Facebook, comparing those published by major mainstream news publishers and those posted by "hoax sites and hyperpartisan blogs".

Other news sources also reported on different aspects of this emerging issue. The Washington Post reported on the economics of fake news websites [6], and the New York Times reported a case study of a fake news story going viral [7]. BuzzFeed also reported on "more than 100 pro-Trump websites being run from a single town in the former Yugoslav Republic of Macedonia" [8]. In response to this increased attention, Facebook and Google both announced that they would start to ban fake news websites from their advertising platforms, and Twitter announced improvements to its reporting functionality.

These issues have not been investigated in this thesis, as they became prominent after the conclusion of the study in Chapter 6. However, they do raise additional questions which should be investigated in future work in this area. Though the position taken in this thesis, and suggested by the relevant literature in Chapter 2, has been that openness in news media is good for democratic society, it is clear from these reports that there are dangers associated with the opening of the news process.

Future research should investigate to what extent these failures of alternative news systems are different from the failures of traditional news media (see Chapter 2), and how these risks can be mitigated in a system where the level of contribution is too high for human moderators to control. Techniques such as those proposed in this document may be useful in ensuring that diverse viewpoints are heard, fake news stories are prevented from spreading, and filter bubbles are broken.

7.6 Conclusions

This thesis has investigated the changing relationship between news producers and consumers, and investigated the effects that changing commenting systems have on the way that people interact with the news.

It found that despite the claims of news outlets, citizens are being given very limited input into the news process, and the problems identified by Gans (2011) and others are

[5] See `https://www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook`. Accessed November 2016

[6] See `https://www.washingtonpost.com/news/the-intersect/wp/2016/11/18/this-is-how-the-internets-fake-news-writers-make-money/`. Accessed November 2016

[7] See `http://www.nytimes.com/2016/11/20/business/media/how-fake-news-spreads.html`. Accessed November 2016

[8] See `https://www.buzzfeed.com/craigsilverman/how-macedonia-became-a-global-hub-for-pro-trump-misinfo`. Accessed November 2016

not being fully addressed. It also found that the systems that do involve citizens more fully in the process are considered to be less credible than those which do not.

Despite identifying an algorithm that increased the diversity of viewpoints shown to users, confirming with users that credible news represents an increased variety of viewpoints, and specifically asking users to view the news article and comments as a combined story, participants still largely focused on the journalist-written article and neglected the comments. This suggests that news readers generally do value the writing of journalists and suggests that for open news systems to become credible is going to take a large shift in what people consider to be news.

For news organisations it does provide permission to experiment. It shows that news consumers can clearly separate the work of journalists (which they use to judge the news outlet) from other contributions (which they do not). This suggests that news outlets can continue to experiment with the ways in which they accept contribution from the public as long as that contribution is clearly separated from the primary journalistic content.

With the usage of non-traditional news sources growing, it is important to understand how they impact on readers' understanding of the world and how the voices of many more people can be included in a responsible way. A population exposed to a greater diversity of viewpoints will be less likely to suffer from the issues identified by Sunstein (2002a) of polarisation and extremism. A media that increases the diversity of viewpoints in its coverage can contribute to multiperspectival news and better fulfil the important tasks required of it in a modern democratic society. It is my hope that the work performed in this thesis can contribute to this effort.

Appendix A

Terms References

Term	**Example Uses**
User Generated Content	Hänska-Ahy and Shapour (2013); Ali and Fahmy (2013); Kperogi (2011); Leung (2009); Hecht and Gergle (2010); Mitchelstein and Boczkowski (2010); Thurman (2008); Hermida and Thurman (2007)
Citizen Journalism	Ali and Fahmy (2013); Bruns and Highfield (2012); Heravi et al. (2012); Väätäjä et al. (2012); Kperogi (2011); Thurman and Hermida (2010); Schlieder and Yanenko (2010); Newman (2009); Bruns (2007); Nip (2006)
Participatory	Loosen and Schmidt (2012); Domingo et al. (2008); Kperogi (2011); Thurman and Hermida (2010); Paulussen and Ugille (2008); Bradshaw (2007); Deuze (2008); Hermida and Thurman (2007)
Social Media	Thurman and Walters (2013); Elmer (2012); Ali and Fahmy (2013); Bruns and Highfield (2012); Hermida et al. (2011); Lerman and Ghosh (2010); Shao (2009)
Filtering / Aggregation / Curation	Thurman and Schifferes (2012); Tang et al. (2011); Peer and Ksiazek (2011); Munson and Resnick (2010)
Collaborative	Väätäjä et al. (2012); Downie and Schudson (2009); Paulussen and Ugille (2008); Papandrea (2006); Hermida and Thurman (2007); Bowman and Willis (2003); Lasica (2003)
Social News	Hermida et al. (2014); Lerman and Ghosh (2010)
Mainstream	Thurman and Hermida (2010); Kwak et al. (2010); Lasica (2003); Newman (2009)
Citizen Reporting	Mustafaraj et al. (2012); Crowley et al. (2012); Starbird and Stamberger (2010); Bruns (2006)
Pro-Am	Downie and Schudson (2009); Bruns (2006)
News Outlet	Diakopoulos et al. (2012); An et al. (2011); Lih (2004)
User Generated Media	Heravi et al. (2012); Shao (2009); Cook et al. (2009)
Crowdsourcing	Daniel and Flew (2010); Väätäjä et al. (2012); Beckett (2010); Newman (2009)
Commentary/Discussion	Johnson and Hyysalo (2012); Bruns and Highfield (2012); Keegan and Gergle (2010)
Microblogging	Elmer (2012); Arceneaux and Weiss (2010); Schlieder and Yanenko (2010)
Alternative	Paulussen and Ugille (2008); Bruns (2006)
Collaborative/Collective Blog	Benkler (2011); Lüders et al. (2010); Haas (2005)
Blog	Malin (2011); Zhou et al. (2011); Bruns (2010); Bowman and Willis (2003)
Personal Publishing	Loosen and Schmidt (2012); Bruns (2007)
Citizen News/Media	Kperogi (2011); Marshall (2007)
Social Media Journalism	Lewis (2012); Pavlik (2013)
Gatewatching	Bruns and Highfield (2012); Bradshaw (2007)
Collaborative Reporting	Väätäjä et al. (2012); Papandrea (2006)
Civic	Kperogi (2011); Kovačič and Erjavec (2008)

Table A.1: A sample of terms used in academic literature

Appendix B

Study of Credibility on The Web Data

Age								
No Answer	**Under 15**	**15-24**	**25-34**	**35-44**	**45-54**	**55-64**	**65+**	
0	0	16	52	3	6	2	0	
Education								
No Answer	**Secondary Education graduated without formal examination qualifications**	**Secondary Education graduated at ordinary or lower examination level**	**Secondary Education graduated at advanced or higher examination level**	**Uncompleted further education College or University**	**Graduate of any further education College or University**	**Masters**	**Doctorate**	
2	1	4	0	13	33	21	5	
Occupation								
No Answer	**Employed For Wages**	**Self Employed**	**Out Of Work**	**Homemaker**	**Student**	**Military**	**Retired**	**Unable To Work**
1	44	4	3	2	24	0	0	1
Gender								
No Answer	**Male**	**Female**	**Other**					
1	57	21	0					

Table B.1: Demographic information for participants in Chapter 4 Credibility study

System	Unbiased					Fair					Whole Story					Accurate					Trustworthy				
	4	**3**	**2**	**1**	**0**	**4**	**3**	**2**	**1**	**0**	**4**	**3**	**2**	**1**	**0**	**4**	**3**	**2**	**1**	**0**	**4**	**3**	**2**	**1**	**0**
BBC News on Facebook	3	17	7	3	2	3	19	7	1	2	0	15	11	5	1	6	16	7	2	1	6	14	7	3	2
BBC News	7	13	9	4	2	10	13	7	5	0	4	17	7	7	0	13	13	6	3	0	10	14	7	4	0
Gawker	0	3	1	3	0	0	2	5	0	0	0	3	0	2	2	1	1	4	0	1	0	1	4	1	1
MP Expenses investigation	1	4	3	0	1	2	5	1	0	1	2	3	2	1	1	2	4	2	0	1	3	3	2	0	1
Google Blog	5	2	7	6	2	4	8	7	3	0	2	5	9	6	0	5	9	7	1	0	3	9	6	2	2
The Guardian	2	12	5	13	1	2	19	10	2	0	2	15	11	5	0	8	14	8	3	0	5	18	7	3	0
Daily Mail	0	3	4	4	19	0	4	4	6	16	0	3	4	7	16	0	3	6	7	14	0	3	5	6	16
The Sun	0	0	2	8	14	0	0	3	9	12	0	0	1	9	14	0	1	4	6	13	0	0	1	10	13
The Telegraph	0	7	7	10	5	1	9	11	7	1	0	11	8	10	0	2	11	9	5	2	2	8	13	5	1
Republic on Facebook	0	4	4	7	4	1	4	8	6	0	1	2	6	8	2	2	5	6	5	1	1	3	9	5	1
Sky News on Twitter	4	5	6	4	7	2	7	6	5	6	0	7	7	7	5	1	9	7	4	5	1	7	10	2	6
Wikileaks	5	10	5	3	1	4	9	8	2	1	5	7	3	6	3	6	12	5	1	0	4	11	8	0	1
Guido Fawkes	1	1	0	3	3	1	3	0	3	1	2	0	3	1	2	1	1	5	0	1	1	2	2	2	1
Slashdot	0	6	3	2	0	1	6	4	0	0	0	4	4	3	0	1	6	3	1	0	1	5	4	1	0
Huffington Post	0	7	7	8	6	0	10	10	5	3	0	9	6	8	5	0	10	7	8	3	0	11	7	6	4
Comment Is Free	0	2	4	12	7	2	10	7	5	1	0	5	6	10	4	1	10	10	3	1	0	9	13	2	1
Tumblr	0	1	7	9	8	0	5	10	7	3	1	2	9	6	7	1	3	12	6	3	1	5	10	5	4
Something Awful Forum	1	2	3	12	4	2	11	4	4	1	1	11	2	7	1	3	8	8	2	1	3	8	9	2	0
/r/AMA	1	4	6	7	5	1	9	11	1	1	0	6	7	8	2	1	12	7	3	0	0	8	11	3	1
Flickr	4	5	11	1	2	4	8	11	0	0	3	4	9	6	1	4	7	10	1	1	3	8	11	1	0
Wikinews	3	2	3	1	0	3	3	2	1	0	1	4	3	0	1	1	2	4	2	0	1	3	3	1	1
/r/Politics	0	1	1	9	5	0	4	4	6	2	0	2	4	7	3	0	2	7	5	2	0	1	6	5	4
Facebook	1	1	4	15	14	0	3	13	14	5	1	4	5	12	13	0	3	8	13	11	1	3	8	9	14
Twitter	2	6	7	7	10	2	6	9	11	4	2	3	9	9	9	0	5	17	6	4	1	6	11	9	5
Youtube	2	5	9	10	9	2	7	13	10	3	1	7	11	9	7	1	9	12	9	4	1	8	13	9	4
/r/PoliticalDiscussion	0	1	1	9	5	0	3	2	10	1	0	3	2	7	4	0	4	7	4	1	0	2	7	4	3

Table B.2: Ratings from Chapter 4 Credibility study

Appendix C

Study of Algorithms for Identifying Sentiment Groups in News Communities Questions

Table C.1: Questions used as part of Chapter 5 Sentiment grouping experiment

Question ID	Group	Similar	Dissimilar	Can't Tell
36775	different	22	2	1
36496	different	15	7	2
36794	different	10	6	8
36843	different	7	15	2
36608	different	3	12	9
36392	different	3	12	9
36462	different	24	1	0
36333	different	3	11	10
36556	different	13	9	3
36487	different	17	5	3
36735	different	4	11	10
36424	different	18	5	2
36491	different	8	14	3
36726	different	1	15	9
36786	different	3	16	6
36389	different	13	9	3
36592	different	2	17	6
36386	different	0	23	1

Continued on next page

Table C.1 – continued from previous page

Question ID	Group	Similar	Dissimilar	Can't Tell
36759	different	8	8	9
36394	different	6	12	7
36552	different	3	12	10
36451	different	7	14	4
36501	different	3	10	12
36354	different	9	4	12
36573	different	4	14	6
36477	different	8	8	8
36400	different	16	8	1
36616	different	9	12	4
36713	different	9	7	8
36326	different	17	6	2
36606	different	15	6	4
36550	different	5	13	6
36638	different	22	2	1
36639	different	10	11	3
36674	different	3	20	1
36712	different	14	6	5
36661	different	8	14	3
36696	different	22	0	2
36679	different	15	5	4
36682	different	5	15	4
36654	different	8	10	6
36702	different	5	16	4
36680	different	8	14	3
36545	different	1	23	0
36641	different	1	23	1
36683	different	16	7	2
36828	different	13	9	2
36766	different	0	20	4
36791	different	1	17	6
36777	different	9	8	7
36839	same	15	2	8
36390	same	12	9	4
36844	same	10	7	7
36810	same	16	5	4

Continued on next page

Table C.1 – continued from previous page

Question ID	Group	Similar	Dissimilar	Can't Tell
36607	same	16	5	3
36489	same	13	6	5
36575	same	11	6	8
36739	same	22	2	1
36750	same	23	1	0
36469	same	12	4	8
36763	same	13	8	4
36541	same	19	4	2
36484	same	22	1	2
36585	same	20	4	1
36734	same	18	4	2
36798	same	11	9	4
36610	same	13	5	6
36519	same	19	4	2
36706	same	1	23	0
36772	same	11	9	5
36586	same	16	3	6
36498	same	16	2	7
36404	same	15	5	4
36378	same	14	7	3
36396	same	12	4	8
36449	same	6	16	2
36434	same	10	8	7
36357	same	8	5	11
36422	same	4	11	10
36350	same	11	7	7
36395	same	14	6	4
36348	same	10	9	5
36825	same	21	2	2
36472	same	5	12	7
36596	same	5	15	4
36414	same	18	3	4
36398	same	9	5	10
36480	same	8	11	6
36643	same	15	7	3
36600	same	5	10	9

Continued on next page

Table C.1 – continued from previous page

Question ID	Group	Similar	Dissimilar	Can't Tell
36441	same	8	15	2
36583	same	17	7	1
36637	same	20	3	1
36755	same	23	1	1
36538	same	23	1	0
36662	same	14	7	3
36437	same	10	7	7
36678	same	9	12	4
36685	same	20	1	3
36681	same	21	3	1

Appendix D

Interface Design

The design process took place over three iterations, each with two sessions of three participants (eighteen participants in total).

Prior to the first session, three basic designs were produced. The first (see Figure D.1) was based entirely on existing online discussion systems, presenting the discussion as a vote-ordered list of nested comments. This was produced so as to provide a basic interface with which the more sophisticated interfaces could be compared.

The second design (see Figure D.2) was a "groups first" approach which included a summary showing a representative comment from each group and allowing users to click on a group to view the users and comments within that group. Due to space contraints only the largest groups were displayed on the menu, however the design also featured an "other" section to display users and contributions that have not been place into one of the largest groups.

The third design (see Figure D.3) was again based on existing online discussion systems but indicated through colour the author's group for each comment. It also allowed readers to click on the group to see more details about that group in a slide-out panel to the right. Instead of ordering the posts chronologically or by popularity, they were instead ordered to maximise the number of groups visible. This is achieved by first showing the most popular post from the largest group, then the most popular from the second largest, etc. until every group has been displayed. Then the system will show the second most popular post from the largest group, then the second largest, and so on until every comment has been shown.

In each iteration, two sets of three participants were given printed versions of the interfaces and also a printed version of an existing news interface showing the same discussion. They were asked to think through how they would interact with the system and to annotate parts of the interface they would change. In the second and third iteration they were also able to interact with and modify a working prototype of each design. Notes

were taken based on their responses and changes made before the next iteration. This section collates the feedback from these sessions and details the changes to the system made in response to this feedback.

D.0.1 Iteration 1

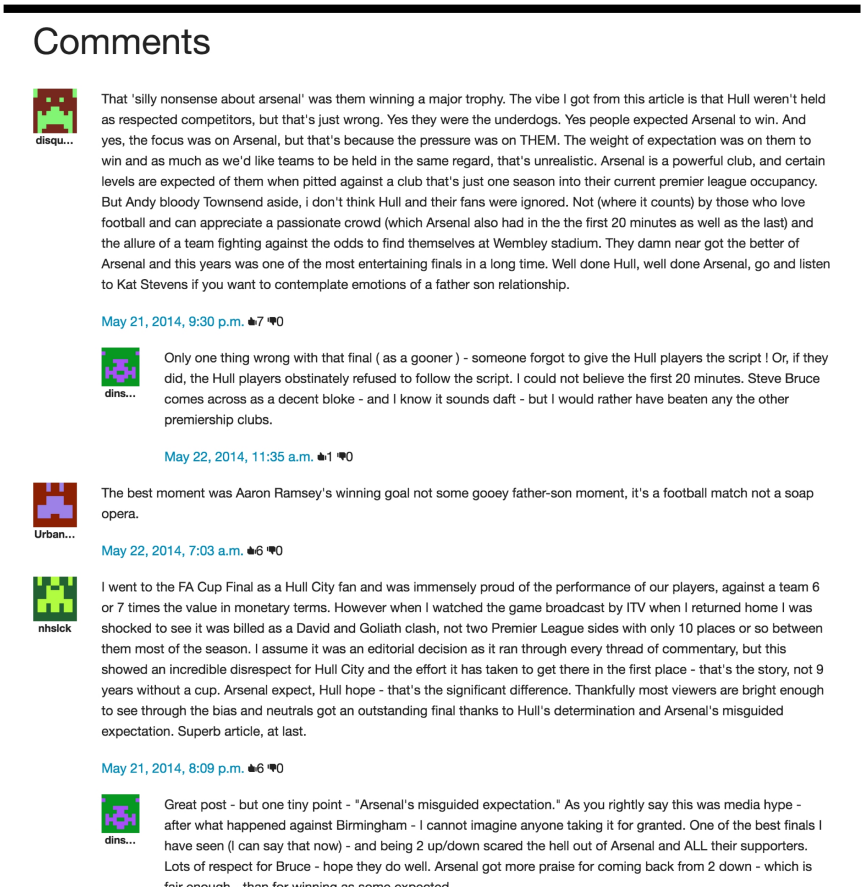


Figure D.1: Iteration 1 Design 1

The participants believed that Design 1 (Figure D.1) captured all of the functionality of the existing interface, except that it was missing contextual dates (e.g. “3 weeks ago”), didn’t have explicit indications of replies (using arrows), and it showed individual vote counts (the number of likes and dislikes) rather than a simple aggregate vote count as in the existing design.

They found that cropping long usernames made them difficult to read, and that users should be able to change the sorting algorithm (by chronology or popularity). They were also distracted by the fact that actual display pictures hadn’t yet been used in the design and temporary placeholders were being used.

Figure D.2: Iteration 1 Design 2

For Design 2 (Figure D.2), participants were initially confused about what the "other" section signified. They thought it implied that the contributions within it were similar, whereas they were actually unrelated. This was especially problematic due to the "representative post" on the summary, which implied that the posts within that group would be similar. Once it was explained, the participants also felt it was confusing that some groups were missing from the summary and only appeared on the "other" section.

One suggested change for the "other" section was to order it chronologically rather than by votes, as since the posts aren't related they aren't based on votes from the same users. However, through discussion, an idea emerged of having a "show all posts" section instead of the "other" section which would not have the same implications. It was decided that this "all posts" section should order comments using the same method as in Design 3 (alternating groups).

Though it was confusing that not all groups were visible on the summary, the participants were also concerned about what would happen at a larger scale if more groups were present. They were concerned that this may make the system difficult to read. It was decided to move the summary to the side of the screen so that they are always visible for context, and so that there is enough space to display a greater number of groups.

They believed that distinguishing groups based only on colour was poor for accessibility, and that the colour contrast on some elements (e.g. links) was too low. It was agreed that the contrast should be increased and the group name should be listed alongside the username as well as the colour being used.

The participants questioned if the groups could be named rather than numbered to aid memory, however it was decided that generating names based on content was both outside the scope of the study and also risks influencing users.

Additional minor suggestions were also made, such as the addition of a clear logical statement describing the filtering (e.g. "showing all posts by Group 1") and listing the amount of comments within a group on the summary as well as the amount of people.

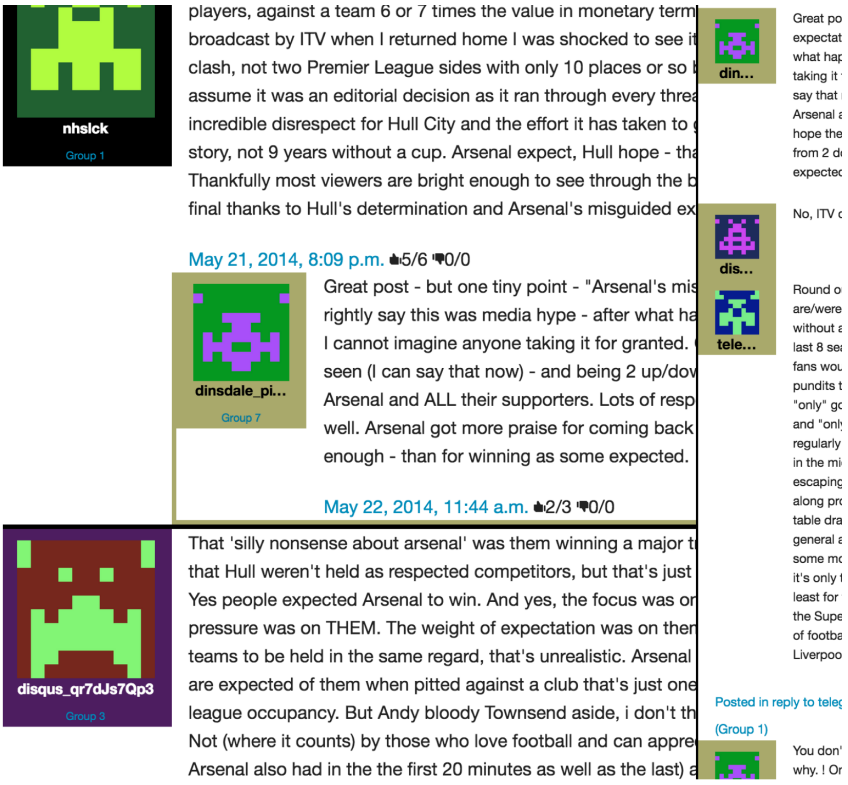


Figure D.3: Iteration 1 Design 3

For Design 3 (Figure D.3) the biggest concern was that while the default sorting behaviour (optimising the number of groups visible) would potentially be useful, the behaviour was not easy to understand. It was agreed that adding options to sort "by group" "by popularity" or "by time" would make this clearer. Participants also felt that as chronological ordering was not used at any level - reading the discussion may become difficult if users do not reply at the correct level (so that the reply appears beside the post they are responding to rather than beneath it).

As this design does not present the group summary, participants believed they would be unable to understand what the colours meant without an explanation. However a solution to this was not agreed in this session.

One participant questioned the stability of the groups, wondering if it will be difficult to follow an ongoing discussion if users are frequently moving groups. The system was modified to ensure that members don't move groups frequently by having colours "stick" to the most central member of the group

There was also a minor design issue in that there wasn't enough whitespace separating unrelated comments.

D.0.2 Iteration 2

After making the changes proposed from the previous sessions. The designs were then placed in front of another two groups of users with experience of online news discussions.

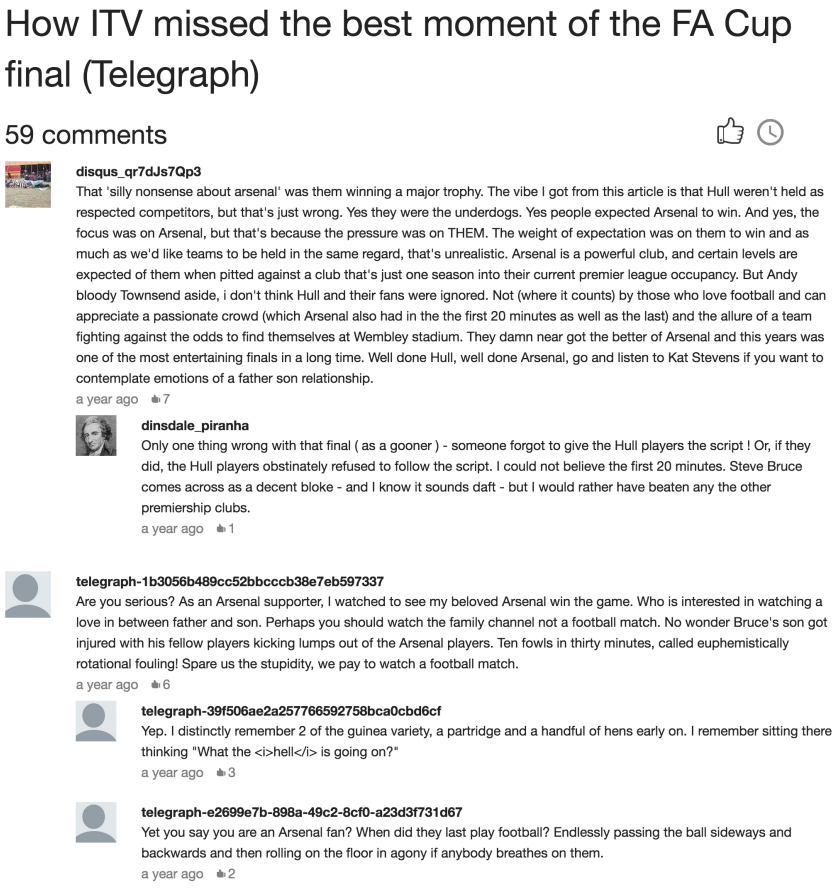


Figure D.4: Iteration 2 Design 1

In this iteration, the participants were satisfied that Design 1 (see Figure D.4) accurately represented existing online discussion systems, and discussion focused on Design 2 and Design 3.

Through discussion the participants agreed that Design 2 was easier to understand and better presented the groups than Design 3 and so it was decided to continue development of Design 2 and to cease work on Design 3.

Figure D.5: Iteration 2 Design 2

The groups looking at the second iteration of Design 2 (see Figure D.5) decided that the group summary was clearer when it was on the top of the page, and that with a large number of groups the summary becomes difficult to understand regardless of its position on the page.

It was felt that the "other" section should be restored, although the filtering statement should make it clear that these are users who do not have a group, they should not have a "representative comment", and they should not be coloured.

It was decided to move the summary back to the top and limit the system to the largest seven groups, leaving all other users ungrouped, which (when adding the "all posts" and "other" sections) ensures that there will be no more than three rows of groups.

The group colours were faded to make them less distracting as they were too prominent and could be difficult to understand when a reader is presented with a page full of bright colours.

The user list was viewed as being too large on all groups, and especially on the "all posts" section where it lists every participant in the discussion. Participants also wanted it to be ordered in a meaningful way. It was decided to order them by a measure of "centrality" (the strength of ties to their group when compared to the strength to other groups), remove the user list entirely from the "all posts" section, and limit it to two rows in all other sections (with text saying "x more" to indicate the total group size).

D.0.3 Iteration 3

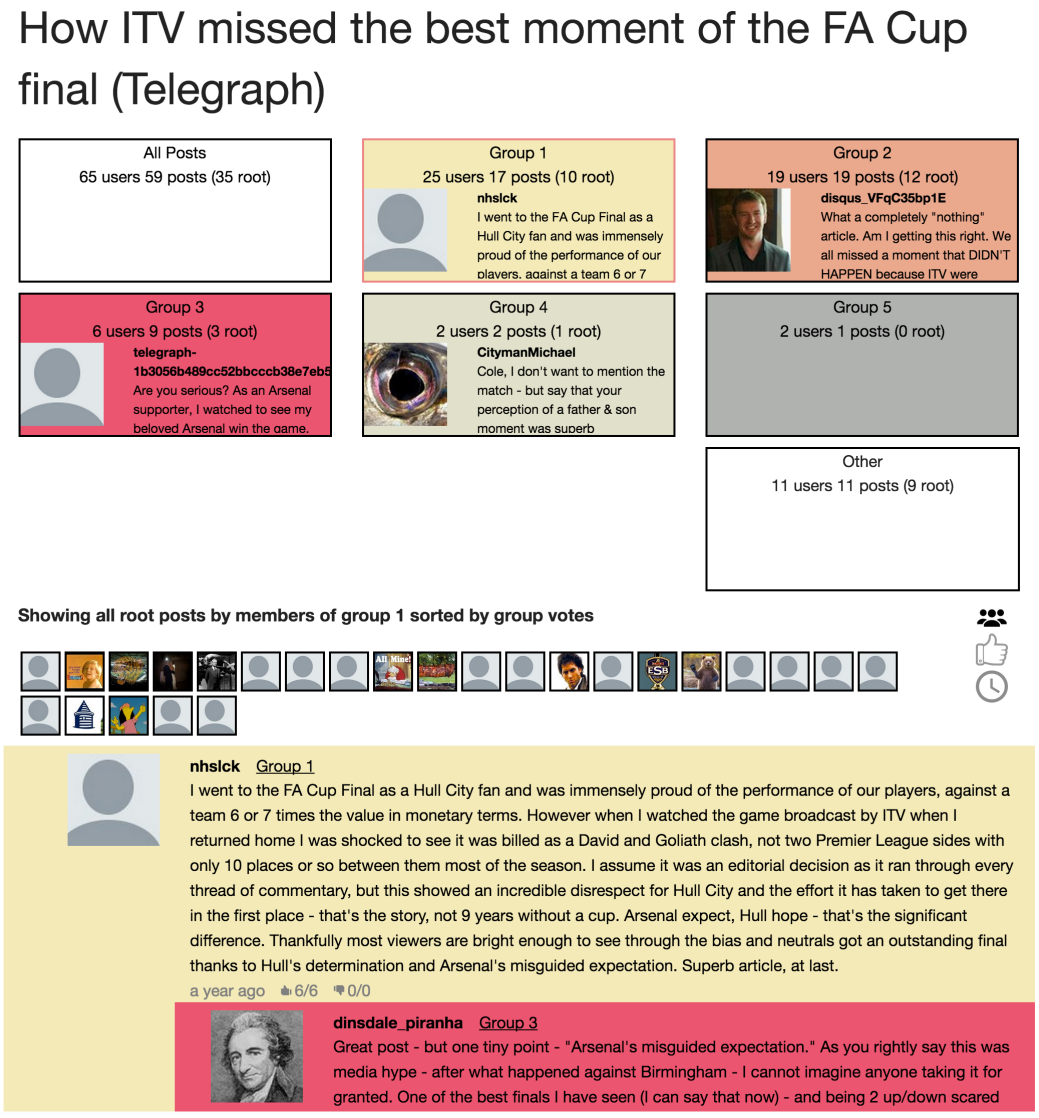


Figure D.6: Iteration 3 Design

In the final iteration, as the re-implementation of the existing interface had been completed, and Design 3 had been abandoned in favour of concentrating on Design 2, the participants only had a single design to improve (see Figure D.6).

The wording used when clicking a summary with no posts to show was considered to be unclear. The users wondered why a group was showing if they had made no comments. This wording was changed to make it clear that members of the group have made replies but have made no "root" comments.

This issue with some groups lacking "root" comments also caused the summary to look inconsistent as these groups were presented without a representative comment. To correct this placeholder content was added for these groups.

There was also some minor issues with the level of indentation and the way that we indicated a post was by an ungrouped member. The identation level was decreased to allow for more comments on screen at once, and the fact that a member was not in a group was made more explicit.

This iteration had fewer comments and minor changes and so it was decided that after making the small changes requested, the final iteration would be produced. See the final design in Figure D.7.

The final design places the range of viewpoints in a very prominent position at the top of the page, where users will be required to see it before seeing the comments. This should help them understand the colour-coded comments that they see further down the page, and also ensures that the key change to the system (the addition of these groups) will influence the readers' perception (consistent with Fogg's Prominence-Interpretation Theory). This understanding of the groups should also be enhanced by the presence of the three options for sorting contribution (order by groups, popularity, or chronology) as this makes it explicit what is happening.

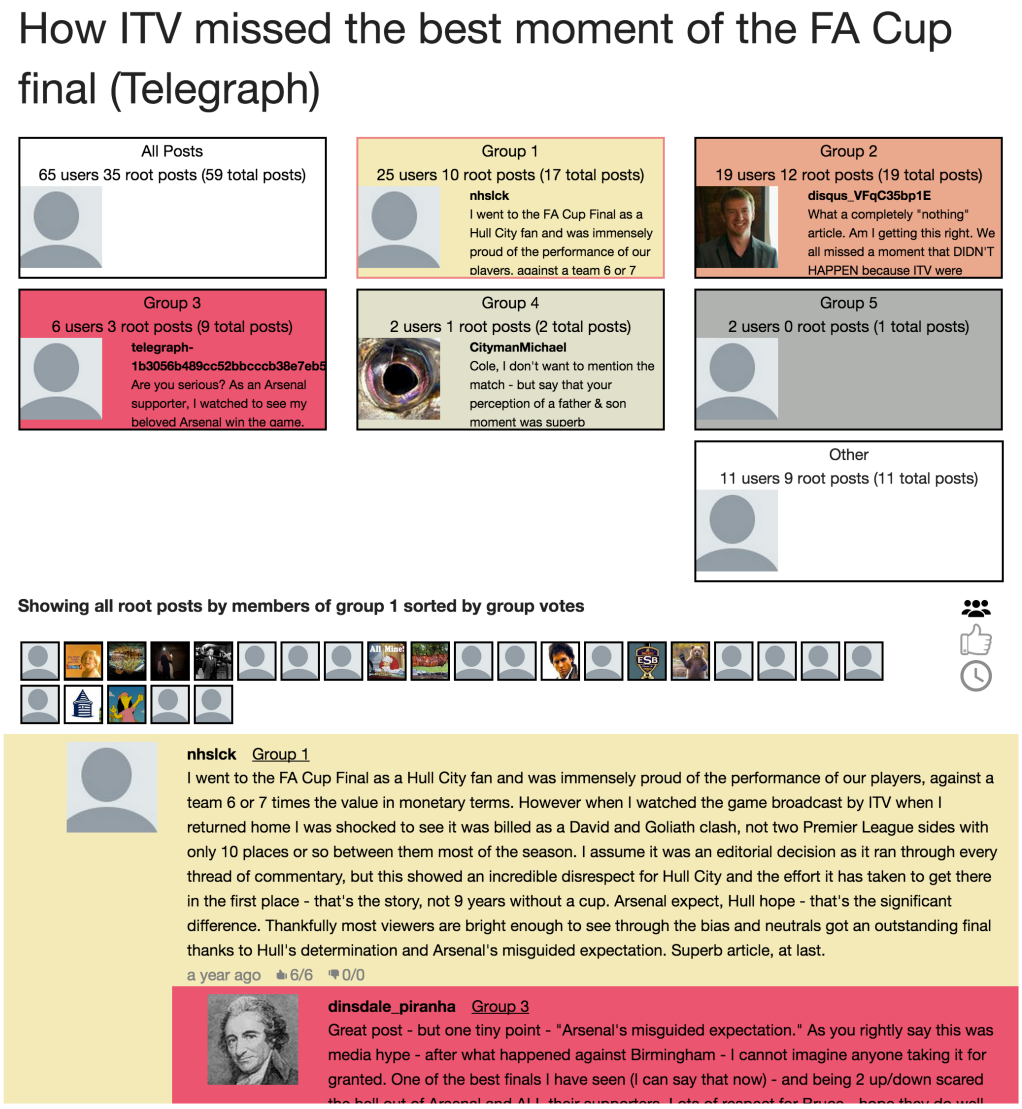


Figure D.7: Final Design

Appendix E

Experimental validation notes

E.1 Participant 1

This participant used the new interface designed to best present the new structure. The page also included the original news article.

Tutorial

- Part of the "tutorial" text was cut off by default (scrolled off the top of the screen). Due to the styling the participant didn't notice and only read the second part.
- The participant didn't realise what the tabs on the summary were for initially
- The tutorial wording explaining the like and dislike counts was unclear. It should more clearly explain the difference between votes by users in the same group and votes by "all" users.

Story 1

- Spent around 3 minutes reading the story. Continued early with 1:30 left.
- The participant managed to read all of the comments, including those in the "Other" section.
- The participant wasn't sure why there was only 2 posts shown but the summary said "5 total posts".
- The participant repeatedly reworded sentences to make new information fit - which took time
- Spent a little over 3 minutes writing summary of article, 5 minutes summarising comments.

- When judging "accurate" - considered the accuracy of numbers.

Story 2

- Spent just over 5 minutes reading story, before skipping the end to read comments. Continued with 1 minute remaining.
- Did not read "All posts view", immediately clicked on a group.
- Clicked through all of the groups.
- Spent 3 and a half minutes summarising the story. 4 and a half minutes summarising comments.

Story 3

- Spent 3 minutes reading the article. Continued with 2 minutes remaining.
- Clicked through every group.
- Spent 2 minutes summarising the story, 7 minutes summarising comments.

E.2 Participant 2

This participant used the "traditional" interface, and was not presented with the story.

Tutorial

- This tutorial was much smaller due to the absence of groups and the participant completed it a lot quicker.

Story 1

- The pariticpant ran out of time with around 3/4 of the comments read.
- The participant spent 1 minute summarising the key points from the comments.

Story 2

- The participant ran out of time with around 1/4 of the comments read.
- The participant spent 1 minute summarising the key points from the comments.

Story 3

- The participant ran out of time with around 1/4 of the comments read.
- The participant spent 1 minute summarising the key points from the comments.

Appendix F

Consuming Experiment Demographics

Age								
No Answer	**Under 15**	**15-24**	**25-34**	**35-44**	**45-54**	**55-64**	**65+**	
0	0	10	37	9	5	2	0	
Education								
No Answer	**Secondary Education graduated without formal examination qualifications**	**Secondary Education graduated at ordinary or lower examination level**	**Secondary Education graduated at advanced or higher examination level**	**Uncompleted further education College or University**	**Graduate of any further education College or University**	**Masters**	**Doctorate**	
1	0	1	1	5	17	32	6	
Occupation								
No Answer	**Employed For Wages**	**Self Employed**	**Out Of Work**	**Homemaker**	**Student**	**Military**	**Retired**	**Unable To Work**
0	28	5	0	1	29	0	0	0
Gender								
No Answer	**Male**	**Female**	**Other**					
0	31	32	0					

Table F.1: Demographic information for participants in Chapter 4 Credibility study

Appendix G

Consuming Experiment Credibility Results

	Old I Old A (42)	**Old I New A (51)**	**New I Old A (51)**	**New I New A (42)**	**Old I (93)**	**New I (93)**	**Old A (93)**	**New A (93)**
Fair	**2.452 (σ=1.138)**	2.353 (σ=0.946)	2.039 (σ=1.009)	2.238 (σ=1.171)	**2.398 (σ=1.038)**	2.129 (σ=1.090)	2.226 (σ=1.089)	**2.301 (σ=1.055)**
Unbiased	2.000 (σ=1.397)	**2.157 (σ=1.194)**	1.804 (σ=1.121)	1.643 (σ=1.250)	**2.086 (σ=1.292)**	1.731 (σ=1.184)	1.892 (σ=1.257)	**1.925 (σ=1.246)**
Wholestory	1.690 (σ=1.439)	**2.000 (σ=1.283)**	1.608 (σ=1.173)	1.929 (σ=1.316)	**1.860 (σ=1.365)**	1.753 (σ=1.250)	1.645 (σ=1.300)	**1.968 (σ=1.299)**
Accurate	2.310 (σ=1.102)	**2.451 (σ=1.016)**	2.039 (σ=1.084)	2.238 (σ=1.019)	**2.387 (σ=1.058)**	2.129 (σ=1.060)	2.161 (σ=1.100)	**2.355 (σ=1.023)**
Trust	2.024 (σ=1.205)	**2.451 (σ=0.976)**	1.941 (σ=1.110)	2.071 (σ=1.242)	**2.258 (σ=1.106)**	2.000 (σ=1.173)	1.978 (σ=1.155)	**2.280 (σ=1.120)**

Table G.1: All average ratings

	Four Conditions	**Interface**	**Algorithm**
Fair	H = 3.346 (p >0.05)	U = 4902.000 (p >0.05)	U = 4226.000 (p >0.05)
Unbiased	H = 4.539 (p >0.05)	U = 4990.500 (p >0.05)	U = 4240.000 (p >0.05)
Wholestory	H = 2.736 (p >0.05)	U = 4474.000 (p >0.05)	U = 3741.000 (p >0.05)
Accurate	H = 2.973 (p >0.05)	U = 4848.500 (p >0.05)	U = 3964.500 (p >0.05)
Trust	H = 5.167 (p >0.05)	U = 4820.000 (p >0.05)	U = 3752.500 (p >0.05)

Table G.2: All Kruskal Wallis H and Mann-Whitney U

	Old I Old A (14)	**Old I New A (17)**	**New I Old A (17)**	**New I New A (14)**	**Old I (31)**	**New I (31)**	**Old A (31)**	**New A (31)**
Fair	**2.571 (σ=1.116)**	2.176 (σ=0.984)	1.941 (σ=0.998)	2.429 (σ=1.116)	**2.355 (σ=1.064)**	2.161 (σ=1.081)	2.226 (σ=1.099)	**2.290 (σ=1.053)**
Unbiased	**2.000 (σ=1.414)**	1.882 (σ=1.131)	1.824 (σ=1.200)	1.571 (σ=1.116)	**1.935 (σ=1.268)**	1.710 (σ=1.169)	**1.903 (σ=1.304)**	1.742 (σ=1.135)
Wholestory	2.214 (σ=1.319)	2.118 (σ=1.131)	1.824 (σ=1.097)	**2.357 (σ=0.972)**	**2.161 (σ=1.221)**	2.065 (σ=1.076)	2.000 (σ=1.218)	**2.226 (σ=1.069)**
Accurate	**2.500 (σ=0.824)**	2.235 (σ=1.214)	2.118 (σ=0.963)	2.286 (σ=0.795)	**2.355 (σ=1.064)**	2.194 (σ=0.895)	**2.290 (σ=0.923)**	2.258 (σ=1.046)
Trust	**2.429 (σ=1.116)**	2.412 (σ=1.088)	2.059 (σ=0.998)	2.071 (σ=1.223)	**2.419 (σ=1.101)**	2.065 (σ=1.105)	2.226 (σ=1.069)	**2.258 (σ=1.163)**

Table G.3: Story 1 average ratings

	Four Conditions	**Interface**	**Algorithm**
Fair	H = 2.887 (p >0.05)	U = 530.000 (p >0.05)	U = 473.500 (p >0.05)
Unbiased	H = 0.673 (p >0.05)	U = 525.500 (p >0.05)	U = 505.000 (p >0.05)
Wholestory	H = 1.777 (p >0.05)	U = 497.500 (p >0.05)	U = 434.500 (p >0.05)
Accurate	H = 0.631 (p >0.05)	U = 515.000 (p >0.05)	U = 497.500 (p >0.05)
Trust	H = 1.680 (p >0.05)	U = 569.000 (p >0.05)	U = 474.000 (p >0.05)

Table G.4: Story 1 Kruskal Wallis H and Mann-Whitney U

	Old I Old A (14)	**Old I New A (17)**	**New I Old A (17)**	**New I New A (14)**	**Old I (31)**	**New I (31)**	**Old A (31)**	**New A (31)**
Fair	2.500 (σ=1.239)	**2.529 (σ=0.915)**	2.059 (σ=0.937)	2.429 (σ=1.178)	**2.516 (σ=1.074)**	2.226 (σ=1.069)	2.258 (σ=1.106)	**2.484 (σ=1.043)**
Unbiased	2.357 (σ=1.394)	**2.412 (σ=1.141)**	1.941 (σ=1.056)	1.857 (σ=1.407)	**2.387 (σ=1.262)**	1.903 (σ=1.228)	2.129 (σ=1.238)	**2.161 (σ=1.298)**
Wholestory	1.214 (σ=1.372)	**1.882 (σ=1.323)**	1.294 (σ=0.956)	1.786 (σ=1.473)	**1.581 (σ=1.386)**	1.516 (σ=1.241)	1.258 (σ=1.163)	**1.839 (σ=1.393)**
Accurate	2.286 (σ=1.097)	**2.588 (σ=0.911)**	2.176 (σ=1.042)	2.286 (σ=1.030)	**2.452 (σ=1.011)**	2.226 (σ=1.038)	2.226 (σ=1.069)	**2.452 (σ=0.978)**
Trust	2.000 (σ=1.195)	**2.471 (σ=0.915)**	1.941 (σ=1.056)	2.214 (σ=1.264)	**2.258 (σ=1.077)**	2.065 (σ=1.162)	1.968 (σ=1.121)	**2.355 (σ=1.094)**

Table G.5: Story 2 average ratings

	Four Conditions	**Interface**	**Algorithm**
Fair	H = 1.855 (p >0.05)	U = 555.000 (p >0.05)	U = 432.000 (p >0.05)
Unbiased	H = 2.308 (p >0.05)	U = 584.500 (p >0.05)	U = 473.000 (p >0.05)
Wholestory	H = 2.884 (p >0.05)	U = 486.500 (p >0.05)	U = 370.500 (p >0.05)
Accurate	H = 1.374 (p >0.05)	U = 537.500 (p >0.05)	U = 426.000 (p >0.05)
Trust	H = 2.142 (p >0.05)	U = 528.000 (p >0.05)	U = 392.000 (p >0.05)

Table G.6: Story 2 Kruskal Wallis H and Mann-Whitney U

	Old I Old A (14)	**Old I New A (17)**	**New I Old A (17)**	**New I New A (14)**	**Old I (31)**	**New I (31)**	**Old A (31)**	**New A (31)**
Fair	2.286 (σ=1.030)	**2.353 (σ=0.904)**	2.118 (σ=1.078)	1.857 (σ=1.125)	**2.323 (σ=0.963)**	2.000 (σ=1.107)	**2.194 (σ=1.060)**	2.129 (σ=1.039)
Unbiased	1.643 (σ=1.288)	**2.176 (σ=1.248)**	1.647 (σ=1.081)	1.500 (σ=1.180)	**1.935 (σ=1.294)**	1.581 (σ=1.129)	1.645 (σ=1.179)	**1.871 (σ=1.263)**
Wholestory	1.643 (σ=1.445)	**2.000 (σ=1.372)**	1.706 (σ=1.362)	1.643 (σ=1.342)	**1.839 (σ=1.416)**	1.677 (σ=1.353)	1.677 (σ=1.400)	**1.839 (σ=1.370)**
Accurate	2.143 (σ=1.301)	**2.529 (σ=0.848)**	1.824 (σ=1.200)	2.143 (σ=1.187)	**2.355 (σ=1.094)**	1.968 (σ=1.204)	1.968 (σ=1.257)	**2.355 (σ=1.033)**
Trust	1.643 (σ=1.172)	**2.471 (σ=0.915)**	1.824 (σ=1.248)	1.929 (σ=1.223)	**2.097 (σ=1.117)**	1.871 (σ=1.238)	1.742 (σ=1.217)	**2.226 (σ=1.099)**

Table G.7: Story 3 average ratings

	Four Conditions	**Interface**	**Algorithm**
Fair	H = 1.565 (p >0.05)	U = 547.000 (p >0.05)	U = 504.500 (p >0.05)
Unbiased	H = 2.870 (p >0.05)	U = 548.500 (p >0.05)	U = 429.500 (p >0.05)
Wholestory	H = 0.777 (p >0.05)	U = 510.500 (p >0.05)	U = 446.500 (p >0.05)
Accurate	H = 2.596 (p >0.05)	U = 563.000 (p >0.05)	U = 399.500 (p >0.05)
Trust	H = 4.361 (p >0.05)	U = 509.500 (p >0.05)	U = 382.500 (p >0.05)

Table G.8: Story 3 Kruskal Wallis H and Mann-Whitney U

Appendix H

Consuming Experiment Credibility Reasons

	Old I Old A	**Old I New A**	**New I Old A**	**New I New A**	**Old I**	**New I**	**Old A**	**New A**
Comments disagree with article	**1**	0	0	0	**1**	0	**1**	0
Mentioned comments	**5**	1	1	**5**	**6**	**6**	**6**	**6**
General lack of trust in news	**2**	0	1	0	**2**	1	**3**	0
Lack of knowledge/information about subject	**3**	0	0	0	**3**	0	**3**	0
Journalist/Editor is biased	**1**	0	**1**	**1**	1	**2**	**2**	1
Sensationalised	0	**1**	0	0	**1**	0	0	**1**
Missing information in article	1	1	**3**	1	2	**4**	**4**	2
Dispute in comments	**1**	0	0	**1**	**1**	**1**	**1**	**1**
Story is simple retelling of facts	**1**	**1**	**1**	0	**2**	1	**2**	1
Mentioned groups	0	0	1	**2**	0	**3**	1	**2**
News outlets are biased	**2**	1	0	0	**3**	0	**2**	1
Mentioned headline	**1**	0	**1**	0	**1**	**1**	**2**	0
Things in comments not mentioned in article	**1**	0	0	0	**1**	0	**1**	0
Lack of perspectives	1	1	**3**	2	2	**5**	**4**	3
Representative of perspectives	1	0	0	**3**	1	**3**	1	**3**
Mentioned article	**6**	4	**6**	4	**10**	**10**	**12**	8
Number of reasons	9.214 (σ=3.707)	8.353 (σ=1.998)	7.647 (σ=3.547)	8.400 (σ=3.262)	8.742 (σ=2.929)	8.000 (σ=3.437)	8.355 (σ=3.703)	8.375 (σ=2.666)

Table H.1: Story 1 Credibility Reasons

	Old I Old A	**Old I New A**	**New I Old A**	**New I New A**	**Old I**	**New I**	**Old A**	**New A**
Missing information in article	**8**	2	1	5	**10**	6	**9**	7
Lack of knowledge/information about subject	**1**	0	0	0	**1**	0	**1**	0
Journalist/Editor is biased	**1**	0	**1**	0	**1**	**1**	**2**	0
Sensationalised	0	1	**2**	0	1	**2**	**2**	1
General lack of trust in news	**1**	0	**1**	0	**1**	**1**	**2**	0
Expectation for news to be true	**3**	1	0	0	**4**	0	**3**	1
Mentioned comments	1	2	1	**4**	3	**5**	2	**6**
Story is simple retelling of facts	1	**2**	0	**2**	**3**	2	1	**4**
Mentioned groups	0	0	0	**2**	0	**2**	0	**2**
Comments disagree with article	0	0	**1**	0	0	**1**	**1**	0
Mentioned headline	1	**2**	0	0	**3**	0	1	**2**
Comments are biased	0	1	0	**3**	1	**3**	0	**4**
Lack of perspectives	**3**	0	0	2	**3**	2	**3**	2
Comments not accurate	1	1	0	**2**	**2**	**2**	1	**3**
Mentioned article	**10**	6	4	6	**16**	10	**14**	12
Number of reasons	11.071 (σ=3.453)	9.529 (σ=2.278)	6.588 (σ=2.302)	8.600 (σ=4.046)	10.226 (σ=2.970)	7.531 (σ=3.391)	8.613 (σ=3.643)	9.094 (σ=3.263)

Table H.2: Story 2 Credibility Reasons

	Old I Old A	**Old I New A**	**New I Old A**	**New I New A**	**Old I**	**New I**	**Old A**	**New A**
Mentioned groups	0	0	**1**	**1**	0	**2**	**1**	**1**
Lack of knowledge/information about subject	**2**	0	0	0	**2**	0	**2**	0
Journalist/Editor is biased	**3**	1	0	**3**	**4**	3	3	**4**
Sensationalised	1	1	**2**	0	**2**	**2**	**3**	1
Mentioned comments	2	1	1	**3**	3	**4**	3	**4**
Expectation for news to be true	1	**2**	0	0	**3**	0	1	**2**
Story is simple retelling of facts	**1**	**1**	0	0	**2**	0	**1**	**1**
Lack of perspectives	**1**	0	0	0	**1**	0	**1**	0
News outlets are biased	**2**	0	0	1	**2**	1	**2**	1
Dispute in comments	0	0	0	**1**	0	**1**	0	**1**
Missing information in article	**2**	1	1	**2**	**3**	**3**	**3**	**3**
Comments are biased	**1**	0	0	0	**1**	0	**1**	0
Things in comments not mentioned in article	0	1	0	**3**	1	**3**	0	**4**
General lack of trust in news	**1**	**1**	0	**1**	**2**	1	1	**2**
Mentioned article	**9**	7	3	6	**16**	9	12	**13**
Mentioned headline	0	0	0	**1**	0	**1**	0	**1**
Number of reasons	8.571 (σ=2.821)	6.588 (σ=2.871)	5.412 (σ=2.402)	7.533 (σ=4.080)	7.484 (σ=3.015)	6.406 (σ=3.463)	6.839 (σ=3.038)	7.031 (σ=3.522)

Table H.3: Story 3 Credibility Reasons

Appendix I

Article and Comment Summaries

	Old I Old A	**Old I New A**	**New I Old A**	**New I New A**	**Old I**	**New I**	**Old A**	**New A**
Success/Awards	10	**13**	7	7	**23**	14	17	**20**
Wasn't Considered Beautiful	12	**16**	15	11	**28**	26	**27**	**27**
Gained first role without audition	**9**	8	5	7	**17**	12	14	**15**
New Film	4	4	**6**	5	8	**11**	**10**	9
Issues with facial features	3	**8**	3	1	**11**	4	6	**9**
Does not consider herself to be beautiful	**6**	5	**6**	4	**11**	10	**12**	9
Never Changed Her Look	3	**5**	**5**	3	**8**	**8**	**8**	**8**
Regarded as being very beautiful	**6**	2	2	1	**8**	3	**8**	3
Shyness	**5**	**5**	4	3	**10**	7	**9**	8
Almost quit acting	2	5	**6**	4	7	**10**	8	**9**
Speaking at Cannes Film Festival	3	3	4	**6**	6	**10**	7	**9**
Number of topics	9.143 (σ=3.642)	8.176 (σ=2.065)	7.176 (σ=3.258)	8.571 (σ=3.310)	8.613 (σ=2.926)	7.806 (σ=3.355)	8.065 (σ=3.574)	8.355 (σ=2.707)

Table I.1: Story 1 Article topics

	Old I Old A	**Old I New A**	**New I Old A**	**New I New A**	**Old I**	**New I**	**Old A**	**New A**
Misogyny	1	3	**4**	1	4	**5**	**5**	4
Discussion of plastic surgery	3	**6**	1	5	**9**	6	4	**11**
Sophia Loren is not attractive	3	8	6	**9**	11	**15**	9	**17**
Plastic Surgery Story	**4**	1	0	3	**5**	3	**4**	**4**
Personality is important	2	**4**	2	2	**6**	4	4	**6**
Reference to "Groups"	0	0	1	**4**	0	**5**	1	**4**
Reference to comment position in discussion	**3**	1	1	1	**4**	2	**4**	2
Sophia Loren is talented	2	**4**	3	2	**6**	5	5	**6**
Reference to number of commenters	7	7	7	**9**	14	**16**	14	**16**
Attention seeking/PR Stunt	1	**3**	1	2	**4**	3	2	**5**
Sophia Loren is not talented	1	1	0	**3**	2	**3**	1	**4**
Sophia Loren is attractive	11	**17**	15	10	**28**	25	26	**27**
Number of topics	9.143 (σ=3.642)	8.176 (σ=2.065)	7.176 (σ=3.258)	8.571 (σ=3.310)	8.613 (σ=2.926)	7.806 (σ=3.355)	8.065 (σ=3.574)	8.355 (σ=2.707)

Table I.2: Story 1 Comments topics

	Old I Old A	**Old I New A**	**New I Old A**	**New I New A**	**Old I**	**New I**	**Old A**	**New A**
Wife filed for divorce	2	**3**	1	1	**5**	2	3	**4**
Wealth	8	**11**	4	5	**19**	9	12	**16**
Landmark ruling	**9**	4	3	8	**13**	11	**12**	**12**
Monaco Football Club	2	2	**5**	2	4	**7**	**7**	4
Ongoing 6 years	5	3	**6**	3	8	**9**	**11**	6
Half will go to wife	13	**17**	13	10	**30**	23	26	**27**
Married for 23 years	5	**7**	4	2	**12**	6	**9**	**9**
Source of Wealth	8	**13**	11	5	**21**	16	**19**	18
Both Sides Claim Victory	**5**	**5**	4	1	**10**	5	**9**	6
Number of topics	11.071 (σ=3.453)	9.294 (σ=2.108)	6.471 (σ=2.329)	8.571 (σ=4.118)	10.097 (σ=2.933)	7.419 (σ=3.425)	8.548 (σ=3.688)	8.968 (σ=3.198)

Table I.3: Story 2 Article topics

	Old I Old A	**Old I New A**	**New I Old A**	**New I New A**	**Old I**	**New I**	**Old A**	**New A**
Money Made Illegitimately	**9**	5	6	**9**	14	**15**	**15**	14
He's so wealthy it doesn't matter	**4**	**4**	**4**	2	**8**	6	**8**	6
Misogyny	**7**	6	5	6	**13**	11	**12**	**12**
Nanny	0	**6**	0	1	**6**	1	0	**7**
Wife Deserves The Money	6	**7**	6	5	**13**	11	**12**	**12**
Marriage Law Favours Women	**6**	2	0	4	**8**	4	**6**	**6**
Wife Just Wanted Money	**7**	4	2	2	**11**	4	**9**	6
Prostitution	0	0	**2**	**2**	0	**4**	**2**	**2**
Reference to demographics of commenters	1	**3**	1	2	**4**	3	2	**5**
Criticism of Russia	**4**	1	2	0	**5**	2	**6**	1
Question the value of marriage	6	**13**	6	10	**19**	16	12	**23**
Prenup	3	**12**	5	4	**15**	9	8	**16**
Reference to number of commenters	**6**	5	5	4	**11**	9	**11**	9
Isa Hussani loves CNN	0	**1**	0	**1**	**1**	**1**	0	**2**
Wife Does Not Deserve The Money	**8**	6	4	5	**14**	9	**12**	11
Number of topics	11.071 (σ=3.453)	9.294 (σ=2.108)	6.471 (σ=2.329)	8.571 (σ=4.118)	10.097 (σ=2.933)	7.419 (σ=3.425)	8.548 (σ=3.688)	8.968 (σ=3.198)

Table I.4: Story 2 Comments topics

	Old I Old A	**Old I New A**	**New I Old A**	**New I New A**	**Old I**	**New I**	**Old A**	**New A**
Hamilton/Rosberg rivalry	**9**	8	7	**9**	**17**	16	16	**17**
Rivalry compared to Senna/Prost	3	**8**	4	4	**11**	8	7	**12**
Hamilton prevented from qualifying in first	**10**	**10**	6	7	**20**	13	16	**17**
Hamilton may think it was intentional	**9**	8	5	7	**17**	12	14	**15**
Rosberg mistake	8	**9**	5	7	**17**	12	13	**16**
Speculation it may have been deliberate	**11**	10	10	8	**21**	18	**21**	18
Previous Schumacher incident	**4**	2	**4**	**4**	6	**8**	**8**	6
Justin Bieber was at Monaco harbour	2	**3**	1	0	**5**	1	**3**	**3**
Formula 1 Monaco Grand Prix	10	**11**	**11**	8	**21**	19	**21**	19
Number of topics	8.357 (σ=2.793)	6.529 (σ=2.913)	5.118 (σ=2.083)	7.714 (σ=4.061)	7.355 (σ=3.001)	6.290 (σ=3.391)	6.581 (σ=2.916)	7.065 (σ=3.528)

Table I.5: Story 3 Article topics

	Old I Old A	**Old I New A**	**New I Old A**	**New I New A**	**Old I**	**New I**	**Old A**	**New A**
Past actions of Hamilton	0	1	0	**2**	1	**2**	0	**3**
It was intentional	10	**11**	9	**11**	**21**	20	19	**22**
Reference to "Groups"	0	0	0	**1**	0	**1**	0	**1**
Hamilton has bad attitude	1	2	**4**	2	3	**6**	**5**	4
Racism	0	1	**3**	0	1	**3**	**3**	1
It was accidental	5	5	**6**	5	10	**11**	**11**	10
Reference to comment position in discussion	0	**1**	**1**	0	**1**	**1**	**1**	**1**
Reference to number of commenters	**7**	4	1	**7**	**11**	8	8	**11**
In-car footage	0	1	0	**3**	1	**3**	0	**4**
Hamilton has good attitude	**2**	0	**2**	1	2	**3**	**4**	1
Number of topics	8.357 (σ=2.793)	6.529 (σ=2.913)	5.118 (σ=2.083)	7.714 (σ=4.061)	7.355 (σ=3.001)	6.290 (σ=3.391)	6.581 (σ=2.916)	7.065 (σ=3.528)

Table I.6: Story 3 Comments topics

Appendix J

Interview Transcripts

J.1 Condition 1 Old Interface Old Algorithm

J.1.1 Participant 1

	Story 1	Story 2	Story 3
Fair	4	1	3
Unbiased	4	1	3
Wholestory	4	2	2
Accurate	3	3	2
Trust	4	2	2

Table J.1: Participant 1 credibility ratings

When deciding on the credibility, this participant looked specifically for quotations and references to third parties to verify that alleged incidents took place. They believed that because the majority of quotations in the Sophia Loren story were from Sophia Loren herself that this detracted from the credibility. They also believed that if they had been informed of the outlets that reported the stories that this would have made them seem more credible.

The participant struggled to read enough of the comments in the time limit to give an informed opinion about the comments. However, of the comments they did read they felt that many were unrelated to and distracted from the actual facts of the story, and also that they were quite biased. They felt that this did not change their opinion about the story as these were simply disregarded. They also did not sense any ordering or organisation to the comments, believing them to be quite random.

They believed that fairness and bias were quite similar, with both reflecting a need to represent "both" sides of a story. They believed that to tell the "whole story", an

article should not get distracted talking about unrelated points, and also that they should include quotations and references to third parties. This need for quotations and third parties was also present in their explanation of "accuracy" and also "trust". The participant made clear a wish to read multiple sources to confirm facts - and believed that some of the comments, through linking to third party sources, helped establish trust in this way.

I: You saw three stories, then summarised the story and the comments, and then rated them in terms of credibility, yes?

P: Yes.

I: What I'm interested in now is how you approached this and what it was you were looking for when you judged the credibility. What factors you thought were important. I have six questions and when I ask them feel free to refer to any of the three stories as examples. Okay?

P: Okay.

I: First of all, when you were deciding if something was credible or not, what elements did you consider? What did you look at and what did you factor in?

P: Well, it was kind of references to incidents that might have happened. For example the story about Hamilton and Nico Rosberg where other people came into the picture, giving references to the Mercedes management giving opinions about the incident. So there was some kind of truth behind that incident because it was being told by other people who felt that something must have happened between them, and something was not right. For me, it was references being pulled from other people who agree with the facts.

I: Okay, so you looked for quotations and third parties and that sort of thing?

P: Yes.

I: Great. On that story you mentioned the Mercedes staff commented on the incident, was there anything specific in the other stories that made you think the story was credible or non-credible?

P: The last story was too small so I don't think that would have much in it because it was about the Russian billionaire. It didn't have much content. The first one, about Sophia Loren I think I'd not say is the best example of something credible because it all came from her, nobody else. The quotations used in the story were mostly from her and no other person. So I wouldn't think that is a good example of something credible. I think the second one [Story 3] was more credible.

I: Okay, so you felt that the fact that the quotations were all from Sophia Loren's point of view, that detracted from the credibility?

P: Yes, because it was just her way of saying things rather than how others see it. It's just her way of seeing things not other people saying things about her.

I: I understand. Did you think there was anything missing from these stories, that if it was present would make you feel it was more credible?

P: If there had been more references from other news agencies, that would have added to the fact that this has happened, rather than depending on individual quotations, which is kind of fair enough. But it'd be more credible if it had been coming from some trusted news agencies.

I: These stories that you read, two of them came from The Telegraph and one from CNN. You didn't know that because it wasn't on the page, but if that was there would that have changed your opinion at all?

P: Yes. Because if it was from BBC or The Telegraph or The Guardian that'd have an extra effect that these are reliable news agencies and I would trust this because they won't fake things just for the attention of their readers.

I: I see you ran out of time to go through all the comments on two of the stories.

P: Yes.

I: So for the third one [Story 2] at least, did the comments change your opinion of the story at all?

P: For the third one [Story 2], there were some comments that were related to the story that happened and some of them were deviating from the thing the news was about. For example, I've noted it in the comments section, some of them were saying things like "people with that much money should not get married because they will have a risk of losing half of their wealth" and some were completely unrelated like "that money did not belong to him and it should belong to the state" or that Putin was helping this guy. Those things were not related to the actual news, it was completely unrelated. Maybe it was right, I have no judgment on that but it was not related to the actual news of the divorce.

I: So do you think that changed your opinion at all or did you disregard it because it's not relevant?

P: I disregarded that, as in terms of the story I read it was news about the split of money during a divorce - it's not related to how he made the wealth.

It's completely unlinked so I disregarded that. The money he made has got nothing to do with his married life.

I: Okay. So again for the comments that you did get a chance to read, did you feel there was any meaningful order or was it just random?

P: Only for one of them I read through them all. I didn't see any order as such, it was quite random. The first comment, if I remember that right, was saying that if someone has lots of money then he does not need to marry. Then there was a chain of replies to that. After that, the next comment was about him making money from the state so he should pay it back to the state. So they are not linked.

I: That's fine. In the comments, did you feel they fairly represented the range of opinions on the topic? or did you feel some viewpoints were missing or one was overrepresented?

P: Yeah I think on the third one[Story 2] they mostly agreed with the billionaire.

I: You think the comments were on his side?

P: Yeah.

I: Okay. There's one final thing I'd like to do. There was five criteria for credibility: Fair, unbiased, tells the whole story, accurate, and trustworthy. I'd like to go through each of those five and talk about what you think they mean. Okay?

P: Okay

I: How did you decide if story was fair? what were you looking for?

P: I was looking for opinions on both sides, not just one side.

I: Okay. How would you decide if a story was biased?

P: I think it's the same thing, it shouldn't have just one side of the story. There should be some opinions that are against it as well. I'm not saying they should be even - like if there are three people for it there should be three people against it - but the content that people are saying is what matters.

I: So do you think bias is similar to fairness?

P: Sort of, yes.

I: Okay. How did you judge if an article is telling the whole story?

P: Well, if it had information describing events that have happened across a span of time. Like, for example in the first story [Story 1], I'm not interesting about all the information about Sophia Loren's past I'm just interested in the information relating to the news. For example, it was okay to know in that news what she felt and what others have said in that span of time. I would expect the full story to just be related to what the news is about.

I: Okay. I'm looking at your answers here and you rated the Sophia Loren story as being fairly reasonable in terms of telling the whole story. Less so in the Hamilton and Russian Oligarch stories. What was it about, say, the Russian Oligarch story, that made you feel it wasn't telling the whole story - whereas the Sophia Loren one was?

P: Because it had individual quotations in the Sophia Loren one, but not in the last one [Story 2].

I: Okay, fair enough. On to accuracy. How did you decide if a story was accurate?

P: Again, if there were quotations that were being said by various people - maybe against it or for it, but saying a thing has happened, then I'd feel it is accurate. And obviously it doesn't have any reference to a news agency, but maybe I'd look for a news agency saying something against it or for it, which means that the thing has happened.

I: So if you were trying to assess accuracy outside of this experiment, you'd be looking for another outlet?

P: Yes. I would always reference another source to get a full picture of it, but for this one I only have one source so I will stick to things in the content, such as the people who's opinions are noted.

I: Finally, trust. How did you decide if you would trust a story?

P: I'd say if there are other stories supporting this. Like I'd search and see if there are other agencies having similar kinds of news, that'd mean there is some truth to that. If I see a news I'd crosscheck that with something else to see if it is accurate.

I: Within this study though you only had access to a single piece of information, how did you decide trust on this? I imagine it was more difficult.

P: That would also come from some of the comments, from people who linked to other sources of information. That tells me something about that news is truth and is not fake.

I: Great. That's all of the questions I have. Is there anything else you wish to say about the stories and comments you read, or about the system you used, or anything else that I haven't brought up?

P: I'd like to say that you should be given more time, 4 minutes is not sufficient to read all of the comments. The last story was okay because it is short but for the others I think you need more time. Second, is that if people aren't comfortable with the topic, for example if they don't know much about Grand Prix, how it works, what is poll position, etc. they will struggle to read it quickly.

I: Thank you for your time.

J.1.2 Participant 2

	Story 1	**Story 2**	**Story 3**
Fair	3	0	2
Unbiased	3	0	1
Wholestory	2	0	1
Accurate	2	1	1
Trust	2	1	2

Table J.2: Participant 2 credibility ratings

This participant felt that at least in some of the stories the journalist appeared to have their own viewpoint on what happened and allowed that to influence the article. They also indicated the news outlet is important in judging credibility, and they'd like information from multiple outlets before making a decision.

They noticed no meaningful order to the comments but preferred ordering by votes as that indicated majority opinions. They also felt that the comments did represent multiple viewpoints and that this enhanced credibility, however this was not enough to offset the perceived bias in the original article.

They felt that fairness and bias were very similar, and for "whole story" they thought of angles of the story which they would expect to see and looked for them in the article and comments. For "accuracy" they made their own analysis of how true the story was. "Trust" was decided based on their assumption of the news source and their own trust towards similar sources.

They also said that they primarily considered the article when judging credibility, though the comments influenced them in a small way.

I: Looking at the first story, about Rosberg and Hamilton. When evaluating the credibility, what elements did you consider?

P: I think it was the content of the story. It was quite long so I didn't finish reading the whole thing but from what I read it seemed to strongly hint that Rosberg has cheated to stop Hamilton from winning, though it doesn't seem to directly accuse him of doing that. Also, with the title "Hamilton hits out at Rosberg as rivalry erupts", it sounds like the author is biased. Also reading the comments below, the majority of people think that Rosberg has cheated, so I think it's biased towards Hamilton.

I: So you've rated this story as non-credible. Were there any elements to the story you found credible?

P: I look for both sides of the story. I think you can tell bias by judging the tone of the article. The article itself seems to think that Hamilton is the victim. So it started off being biased. In the article it also said they were avoiding each other. I don't think that's very objective. It's there to re-enforce the idea that Hamilton is mad because Rosberg stopped him winning.

I: Was there anything that you felt enhanced credibility?

P: I would say the story itself varied in credibility. In terms of the system I think looking at the comments at least, it showed comments from different kinds of opinions. There are some people who are, for example, in the Russian oligarch divorce story, the article for me is really sympathetic to the rich man as the story focuses mainly on money, whereas in the comments most people are saying that the woman doesn't diserve the money. Although there is a small percentage of people supporting the woman. I think if there is any credibility it might be the diversity of views in the comments.

I: What do you think could be added to make the stories more credible?

P: It's like reading any news story online. Sometimes you make your own judgement over which website you visit. If you visit a credible website like BBC you would assume that the information would be true, but if you visit the Daily Mail you'd probably assume that they're not telling the whole story. So if I was to say a news story is credible I would want to check multiple news systems. This is just an individual story for me.

I: How did the comments affect your opinion on the original story?

P: When I read the comments I tend to look for comments that agree with what I was thinking when reading the story. I look for the comments I expected to see. When Sophia Loren was criticised for being unattractive, I looked at the comments and could see why people thought she was attractive and unattractive. I found it interesting why people formed their opinions.

I: Did you feel there was any meaningful order to the comments?

P: No but I found that ordering by votes worked better for me. Because they were highly voted they had more arguments underneath so it's more interesting to see. Whereas if you order by time they tend to be very disorganised, so you can see different opinions next to each other but you don't know if they are major or minor opinions.

I: Did you feel that the comments you read fairly represented the range of views on the topic?

P: Yeah, I think so. Because the views that come to my mind are all in there. Though some of them are slightly different to what I thought of. So I found that the opinions in the comments are wider than I thought of.

I: And so you did perceive multiple viewpoints?

P: Yes, I also came across viewpoints that wouldn't have came to my mind if I hadn't read the comments.

I: And typically would a story with multiple viewpoints be more or less biased or more or less credible?

P: More credible. Because if they allow people to show opinions right or wrong it means they are less controlled by the platform. Even though the article itself might be biased if people in the comments are criticising the article itself it can still turn out to be fair if it reflects what's right and what's wrong rather than what the editor wants you to believe.

I: Finally, I'm interested in how you interpreted the five credibility criteria. So for the Oligarch story you said it wasn't very fair. What did you mean by Fair?

P: I think the article is unfair and biased.

I: So you saw fairness and bias as being quite similar?

P: Yes

I: What about the whole story?

P: It only really focuses on the money. It focuses on him and where his money comes from. It's not on the relationship or on the woman, it didn't say why they wanted divorce.

I: Did the comments change any of that?

P: I think people in the comments didn't realise that it was missing the woman's side of the story. They are basically lead by the article to only think about the number.

I: What were you using to rate accuracy?

P: Accuracy is if it's a true story or a made up story.

I: And you've made your own judgement on that?

P: Yes. I think it might be a bit accurate in terms of the number, his wealth, etc. but apart from that I think there's not much in there.

I: And finally trust

P: I think because I associate the title and content of the article with a news media such as the Daily Mail so some part of it might be true but it's possible that they're not telling the whole story so I won't trust the article.

I: Through a lot of those answers you mentioned the title and the content of the article. Were you taking the comments into account when making these judgements or just the article?

P: Mostly the article itself. I guess the comments on the article, the majority are complaining about the wife not deserving half his fortune. The proportion sticking up for the wife is very small. That's why I also think this contributes to the whole thing being biased and not telling the whole story.

I: I'm interested in knowing how you interpreted those five criteria, and what you were using to judge the credibility of the stories. Do you have anything to add to what you've said so far?

P: I think when considering if something is credible or not, most people when reading an article have already formed their own opinion towards it. If it goes against their opinion they might think it is biased, like me with the oligarch story. I think if they find the story or the comments agree with them they'll find it trustworthy otherwise they'll find it not credible.

J.2 Condition 2 Old Interface New Algorithm

J.2.1 Participant 3

This participant took the potential motivations of the journalist into account, noting that "if you were going to make something up I don't think you'd make that up". They also valued brevity and believed that the facts present would be easily verified. However

	Story 1	**Story 2**	**Story 3**
Fair	2	3	2
Unbiased	2	3	1
Wholestory	1	2	1
Accurate	1	3	2
Trust	2	3	2

Table J.3: Participant 3 credibility ratings

they viewed the writing style, in the Formula One story in particular, as being of poor quality, without a coherent structure, and felt that detracted from credibility.

The participant claimed that due to the type of stories (fairly fact-based articles as opposed to opinion pieces), the comments did not add as much as they otherwise might have. However, for the Formula One story, they lacked background knowledge and struggled to understand the content of the story, and felt that the comments restated facts in a more accessible way.

They did not notice any meaningful order to the comments. They felt they represented a wide range of viewpoints, but perhaps overrepresented some.

Of the divorce story, which other participants viewed as being biased in favour of the husband, this participant did not feel the journalist put forward either side, and discounted the comments as they overrepresented the viewpoint that the wife just wanted money.

They viewed fairness and bias as both relating to giving all perspectives, or attempting to be objective and just stating facts. They also felt whole story was similar, but different in that it's possible to be unbiased but give only a small amount of detail. When deciding accuracy they looked for specific facts that could be judged as accurate or not, and they viewed trust as a combination of these other factors. They did mention the news outlet when discussing trust.

This participant acknowledged that they concentrated on the article when providing the ratings, paying little attention to the comments.

I: You just read three articles. One about Formula One, one about Sophia Loren, and one about the divorce of a Russian oligarch. You also read the comments. What I'm interested in is how you judged the credibility. I have a few questions, you can use any of the three stories for examples as needed. The first question is when you are evaluating the credibility of a story, what elements did you consider?

P: I guess I read it and was like... The first one, the divorce, seemed reasonably credible because it was just reporting what seemed to be fact that you could

fairly easily go and check. If you were going to make something up I don't think you'd make that up, and that one had quotes in. I think it's quite short, and it seems quite objective. If it had of been a political story I might have looked at the comments to get the other side but to be honest on these I don't think they added a lot, certainly in terms of judging the credibility because on that one they were just slightly misogynistic "gold digging", etc. On the second one [Story 3] I didn't really understand the story, because I had no idea what they were talking about or who any of the people were but they mentioned that Justin Beiber was there and that seemed like a random sentence. That seemed to go all over the place and I thought everything was quite one-sided as well.

I: So do you think writing quality came into it as well then?

P: Yeah in terms of the content it seems to be bits from here and there, it didn't seem particularly coherent. The writing quality seems kind of poor, because I couldn't follow it and every sentence was its own paragraph.

I: What in particular enhanced credibility?

P: I think brevity and the fact it was just stating verifiable facts rather than speculation helped. The last one [Story 1] also seemed not in-credible. None of it is really of any consequence so I don't see why they would have made it up unless she really wants to sell this new film, but it didn't seem implausible.

I: And anything that jumped out as a red flag saying this is not credible?

P: There is nothing by itself that really jumped out, I wouldn't say any of these articles were completely made up but maybe it would not a good place to go to try to understand because it seemed muddled more than made up.

I: Was there anything missing, that if it was added would make it more credible to you?

P: In terms of the Sophia Loren story, some examples of roles would be useful. Actually, a photo in this case because I'm not entirely sure what she looks like but if there was a photo where she had a big mouth that I suppose would lend some credibility to the story.

I: You mentioned that you felt the divorce story was credible and was a statement of facts. From other participants I've heard that it was biased because it lacked the wife's perspective. Did you see that at all?

P: Well it doesn't seem like they put forward either side, they've said what a court ruled. I mean they say it's a win but I suppose you ignore what

lawyers say. Then they've said how much money he's got and how he got it. It doesn't make me think ill of the wife.

I: In any case did the comments change your opinion of the story?

P: On the Formula One story I felt that one or two of the comments restated some of the facts in what I think was a slightly more accessible way. So I think I understood what the allegation was slightly better after reading some of the comments. But I mean you're just wading through the comments to try to find something worth reading and then it's like "Sophia Loren is beautiful", "Sophia Loren is ugly", "Sophia Loren had plastic surgery", "Sophia Loren hasn't had plastic surgery". I guess they're all trying to be fairly fact-based articles and I don't know that the comments add a lot. Whereas on opinion pieces I would read the comments because then there seems like a point to them, and the point is that it's opinion. None of the comments seemed to have any new facts.

I: Did you feel there was any meaningful order to the comments?

P: Where they're threaded they are chronological. Apart from that not that I noticed, no.

I: Did you feel they fairly represented the range of viewpoints on the topics?

P: Well they did seem to cover quite a lot of the bases. I think it was all there but maybe on the divorce one there was too many people saying "should have had a prenup".

I: Are any of them over represented?

P: Yeah. Like she's a gold digger.

I: Finally. I'm interested in the five criteria and how you judged each one. In terms of fairness, what were you looking for to say a story is fair.

P: Fair and unbiased are kind of similar. Trying to give both or however many perspectives there are on something, or being quite objective in just stating facts. I guess that's almost the same as wholestory. Not exactly the same, you could be unbiased but still give a small amount of what happened.

I: So you think fairness and biased are very similar?

P: Yeah

I: For whole story, are you able to explain the difference between that and bias and more than you have? For example in the Sophia Loren story was more fair than it told the whole story.

P: Yeah because it didn't seem like it was unfair to anyone. It didn't make me think badly of anyone or think anyone had done anything wrong really. But it wasn't particularly thorough, there wasn't a lot of detail there. It didn't feel like it was the whole story but it didn't seem like wildly biased.

I: So it's about scale is it?

P: Yeah

I: Okay, and for accuracy. You didn't think the Sophia Loren story was accurate but you did think the divorce story was. What were you looking for when judging accuracy?

P: I guess some facts to be accurate or inaccurate, as opposed to... The Sophia Loren story was a longer article but it didn't make a lot of points because it was just human interest.

I: So you felt it was difficult to judge it because there wasn't so many actual facts?

P: Yeah. If it's just her take on things, is it even relevant that it's accurate? does accuracy come in to it? or is that not the point of it, the point of it is to be interesting and to prompt some reflection. So if it wasn't accurate it wouldn't matter.

I: Finally, trust. How would you judge if you would trust a news story or news system.

P: I guess it's a combination of those other things. You want to feel they are not just pushing their own agenda, to the extent that any of us cannot push our own agenda. Or that they're looking at what happened and trying to report that rather than using something that's happened as an excuse to talk about what they always do, like the Daily Mail does. I didn't feel like any of those stories were really doing that though.

I: So in all of those answers you focused primarily on the story and not so much on the comments. Did you feel that the story overwhelmed your credibility view?

P: I think so. I guess the only time I think about the comments is if they were pointing out something that was wrong with the story, but I wouldn't normally think about them too much, because you don't with the comments because they make you upset.

I: In these stories was there any cases where they pointed something out that was missed by the story or something wrong in the story?

P: Not that I saw

I: That's all the questions I have. What I'm interested in is how you judged credibility and how you responded to these stories. I think we covered everything but is there anything you feel should be said that we haven't covered?

P: No

J.2.2 Participant 4

	Story 1	**Story 2**	**Story 3**
Fair	1	3	1
Unbiased	2	3	1
Wholestory	2	3	0
Accurate	1	2	1
Trust	2	3	1

Table J.4: Participant 4 credibility ratings

This participant saw contradictions between comments and article as harming credibility, and felt that the story was more credible when the comments backed up the article. They also thought that stories that left readers with questions were non-credible.

They did not perceive a meaningful order to the comments, and thought that in some cases the comments fairly represented the viewpoints but in others some viewpoints were excluded. They also felt that after reading the top comments the other comments repeated what had already been said.

When discussing fairness they referenced the truthfulness of information in the story and considering multiple points of view, for unbiased they wanted to ensure it did not favour one point of view over others, and when deciding if something tells the whole story they looked for the presence of multiple views. They saw accuracy as a combination of the other factors, and felt that the comments contributed to trust as they helped judge the level of disagreement with the article. They did use the comments when judging credibility by comparing them against the article.

I: You just read three stories, and for each story you gave credibility ratings and summarise the article and the comments. I've got 7 questions here and I'm primarily interested in what you used to judge the credibility: What makes you think something is credible, what makes you think it isn't credible. If you want to give an example, use any of the three stories. The first question is what elements did you consider when you were judging credibility? Take into account the article, comments, system, anything. What did you use to judge?

P: So I know Sophia Loren is very famous and everything but I couldn't remember how she looks. I was thinking why would people say she's ugly and she can't have a career as an actress. Then when I saw the comments everyone was saying that she's beautiful and she's modest. I was thinking how come no-one offered her a job? It's not so convincing a story.

I: So you thought the content of the story didn't feel convincing?

P: Especially also with the comments. Yes. I didn't watch her movies but she said she's shy but then other people say she has a bolder personality.

I: And the other stories? how did you judge the credibility of those? Not so much wether you thought they were credible, but more the individual aspects that you looked at to judge credibility.

P: The formula one story is a bit weird. He said that he didn't intentionally reverse but then other people were saying he's not convincing. I don't follow this sport so there are a lot of questions. And why did they put Justin Beiber's reaction? It's a bit strange

I: When you say other people said it wasn't convincing, who did you mean by other people?

P: I'm not used to reading this kind of news, so I'm not sure who the people are. I know that he said he didn't do it intentionally and people didn't believe him.

I: So by other people, you meant other people in the story not necessarily the comments, right?

P: Yes the story. I didn't ready many of the comments because the story was a bit long. One of the comments was like "Why reverse?"

I: So we've gone over some things there, but specifically what elements of any story did you feel enhanced credibility?

P: I think in the comments of the divorce story, there's some credibility that people said his money came from the fertiliser factory and it was owned by the state. People didn't disagree with the premise of the story.

I: So you thought the basic facts in that story seemed credible because they weren't disputed?

P: Yes

I: And what elements detracted from credibility? For example on the Formula One story you said that it was relatively non-credible.

P: The story is confusing, it still doesn't tell the story. You still have questions which make you want to search more.

I: So you felt it was incomplete?

P: For me, I didn't understand it. If I'm interested I'm going to look for it elsewhere.

I: In any of the stories, did reading the comments change your opinion on the story at all?

P: I think the Sophia Loren story, because they were convincing in many ways that she is not beautiful, they thought she should quit her career or change her face. Since I don't remember what she looks like I thought "well maybe" but then I saw the comments, and maybe they are just her fans so they say she's beautiful, but yeah.

I: Were the comments entirely positive? or was there any negative ones?

P: All positive that I saw.

I: Did you feel there was any meaningful order to the comments?

P: No I didn't.

I: Did you feel the comments fairly represented the range of views on the subject? or did you feel anything was excluded?

P: In the divorce story there was a lot of comments discussing the opposite of the story. Why the woman is rich, or the money is not his money anyway. So you get all this. But the Sophia Loren story everyone was saying she's good, no-one supported the story.

I: So the Sophia Loren story you thought maybe they didn't represent everything? they only had one side, but in the divorce one they did?

P: Yes, in the divorce one there was a variety of comments, even discussing topics outside of the story. But the Sophia Loren story all of the comments seemed to say that she's beautiful. So this is a bit biased.

I: The last thing I want to do is talk about the five credibility criteria and see how you judged them. First, Fairness, what to you is a fair story?

P: In the first story [Story 3] they didn't consider the guy's side of the story, they just said that he did that and Hamilton lost. For the second one [Story 2], it's fair it's a big fortune, it might not be the most expensive but okay, it's fine, the third story [Story 1] is not fair.

I: Why did you think it wasn't fair?

P: Well she's a famous actress and appeared a lot, so it's not fair.

I: Okay, and how did you judge bias?

P: If it was more towards one side of the story it is biased.

I: So to be unbiased you're looking for balance towards different viewpoints?

P: Yes

I: And when deciding if it told the whole story, how did you judge that?

P: It's kind of if they're telling both sides of the story

I: So is that similar to unbiased for you?

P: No, the whole story is more about having more views.

I: So unbiased would be...

P: biased is if they emphasise one side of the story, and whole story is if they have supporting arguments or things to support more opinions.

I: and accuracy, how did you judge if it was accurate?

P: It's like a combination of the other things. If it's fair, and covers the whole story, it's going to be accurate, so it's kind of derived from what we said.

I: And finally trust, how did you decide if you would trust a story?

P: I think the comments affect the trust a lot

I: In what way do the comments affect the trust?

P: It was on the Sophia Loren story because the comments are contradicting the story

I: So that reduced your trust you think?

P: Yes

I: Okay. That's all the questions I have. What I was interested in is how you judged credibility, what you thought of these factors, and generally how you approached the stories. I think we covered everything, but is there anything that you think needs to be said that we haven't said so far.

P: No. I think the first story [Story 3] was a bit complicated as I didn't know the background, but the second [Story 2] and third [Story 1] I moved on from the story before running out of time, taking a bit of everything.

I: Did you read the comments before moving on?

P: I read some of the comments, like the first five and their responses. After that it just seemed to be giving me the same information.

I: On those credibility factors, apart from trust, the rest seemed to be focused on the article itself. When you were rating did you base it primarily on the article or did you also involve the comments?

P: Also the comments, you compare the article with the comments and get the whole story.

I: That's everything then. Thank you.

J.3 Condition 3 New Interface Old Algorithm

J.3.1 Participant 5

	Story 1	**Story 2**	**Story 3**
Fair	3	3	1
Unbiased	3	2	1
Wholestory	3	2	2
Accurate	3	3	1
Trust	1	0	0

Table J.5: Participant 5 credibility ratings

This participant believed that the stories used were aimed primarily for entertainment rather than news, and as such would necessarily have some element of bias.

When reading stories they looked for specific facts and details which they felt lent credibility, and believed that too many opinions (even from different people) resulted in a less credible story. This was the case particularly in the divorce story where the participant felt that the comments section was full of people making their own opinion known though many of the points were not relevant to the story or based on facts from the story. They found it difficult to assess the credibility of the Formula One story due to lacking prior knowledge of Formula One.

Their answers concentrated on the article rather than the comments, and they did not perceive any kind of meaningful ordering to the comments. They attempted to use the tabbed interface but did not notice any difference when clicking between tabs. They also did not have a lot of time to read the comments on two of the stories (about Sophia Loren and Formula One) as they chose to spend more time on the article instead.

They believed that for an article to be fair it should represent "both" sides of a story, and for bias they'd be looking for the presence of facts as well as journalistic opinion. When assessing if they have been told the whole story they thought about unanswered questions. For accuracy they wanted content of an appropriate length that covers just salient topics. This participant did not trust any of the sources as they felt they did not know enough about them. They indicated that knowing who the journalist or news organisation was would have helped with this.

I: Just now you saw three stories, you provided summaries of the story and comments, and then rated them in terms of credibility.

P: Yep

I: What I'm interested in is how you approached this, what you were looking for when judging credibility: what factors you thought were important, what you think could be added to increase credibility, and so on. When you are answering, feel free to use any of the stories as examples.

P: Okay, so it's not about a particular story?

I: No. First of all, when you were deciding if something was credible or not, what elements did you consider?

P: I think it depended on the story. I think the stories that were more sensational, like glamorous stories, had less content to them. The one about Sophia Loren was more factual, the first one [Story 2] was just "this is the more expensive divorce ever". You wouldn't believe that just by looking at the information in there. The one with Sophia Loren was more about her background and things, so I thought that one was more credible. The last one [Story 3] I didn't find interesting, and I thought there was too much going on in that story.

I: Do you mean you were looking for facts?

P: Well, facts as much as feelings. When I read a story I'm more interested in the story and the people, than how much money the divorce was worth or whatever.

I: You said the Hamilton one wasn't particularly credible, the Sophia Loren one was, and the Oligarch story was less than Sophia Loren, yes?

P: Yes.

I: For the Lewis Hamilton one, I understand it's not an area you know a lot about so it is hard to judge, yes?

P: Yeah, and it was all about him cheating but I thought "What evidence do they have to say he's cheating?" I just felt it was somebody trying to make a story out of something. Whereas the one about Sophia Loren talked about her history and her Oscars and her son being the director of a film. It just felt like it had a little bit more to it.

I: Okay. My second question is what elements do you feel enhanced credibility.

P: Probably the details. Factual details rather than just someone's opinion.

I: And what detracted from credibility?

P: Such-and-such said this, such-and-such said that. That makes it less credible to me because then it's just people's opinions.

I: You didn't feel any of the story about Sophia Loren was opinion?

P: Well it of their opinion at the time - when people were saying to her she's not beautiful or her nose is too long, that kind of thing. It was their opinion. She wasn't being ungraceful saying "look at me". She was more saying she wanted to say how she was, and then she got all these parts.

I: So you felt that was more credible?

P: Yes

I: And to detract, that was more people's opinions?

P: After I finished reading the first one [Story 2] I was reading the comments. There was all these people saying derogatory things about this woman that was hardly mentioned in the story, and it made me forget what the actual story was about.

I: Specifically about that story [Story 2] which you thought was non-credible, is there anything that could have been added that would have made it more credible for you?

P: Two sides of the story. It was just a story about him losing money to her because he had it. That's all I felt the story was about.

I: Another angle on the story would have helped?

P: Yeah.

I: Did you read the comments for all of them?

P: No, I ran out of time for comments on the other ones because I concentrated more on the story. The longer the story was the less focus the comments got really.

I: On the one you did get to read the comments on, did the comments change your opinion on the original article?

P: I suppose they did. I forgot what the article was about but I was just getting frustrated at the comments. I was thinking "Who are you to say that?" because there was nothing in the article to base an opinion on.

I: A lot of conjecture?

P: Yeah

I: In the comments you did read - did you feel there was any order to it or was it just a mess?

P: First I just read down the list, the second time I tried to look through the boxes and decide if there was a set of comments I'd rather read - judging by the pictures of the people. I thought "I don't want to read 'she is this' and 'she is that"'. But that didn't really help. It still looked the same.

I: So you didn't feel there was any particular order to the comments?

P: No.

I: Of the comments you read - did you feel they fairly represented the range of views on the topic?

P: No.

I: So you felt they were one-sided?

P: They kind of went on a tangent away from the story and just became "that's just your opinion this is mine" kind of thing.

I: You didn't think they were particularly relevant to the story?

P: No I think they became less relevant to the story the further in they got.

I: Finally I'd like to go through the five credibility criteria and see what you thought of each of them, okay?

P: Okay.

I: First, Fairness. What were you looking for to decide if something was fair?

P: Both sides of the story, and facts.

I: If it had multiple angles on the story, that'd be fair. If it didn't, it's unfair. Yes?

P: Yeah.

I: What about unbiased then?

P: Well you always expect and article to be biased, because those kinds of stories are more entertainment than news. So you'd expect some bias from somewhere. You'd expect there to be some bias but not too much.

I: What do you mean by bias though? What's an example of bias?

P: It'd depend on who is reporting the story. You'd expect whoever is reporting the story is doing it for a reason, like to get you to buy a magazine or something. But I'd expect it to have some fact in it. It'll include their opinion, but I'd expect it to not be just their opinion.

I: You mentioned the person who wrote it. In this case you didn't know who wrote any of those articles, would knowing that information have helped you?

P: Yeah. I think I'd have found it more credible if it'd been BBC news or something.

I: Okay. For "whole story", how did you decide if they were telling you the whole story or not?

P: At the end I would think do I know everything I want to know about that. Like for the first story there was no other elements to the story so I didn't believe that.

I: Because you could think of other questions that weren't answered?

P: Yeah. If it had touched all aspects of the story I might think it's more credible.

I: Okay. How did you decide if you felt something was accurate?

P: By the amount of information. The less information the less accurate.

I: Okay. You mentioned that the Formula one story was quite short.

P: No, that was long. That had too much information in it.

I: So it needs to be the right amount of information?

P: Yeah.

I: And finally trust. How did you decide if you could trust something?

P: I didn't decide I could trust any of them because I didn't know enough about them.

I: You decided not to trust those. Because you didn't have enough information on the source or the actual events?

P: I think the first one [Story 2] the source. The second one [Story 1] I think I trusted more because there was some facts to it. You can't pretend there's a festival going on that hasn't happened. But the last one [Story 3], I thought they were telling me lies.

I: In all of your answers there your answers were pretty much entirely about the story content. Did the comments influence your thinking about the stories at all?

P: I tried not to let them after I read the first one. After I read the first one which was just lots of tangental opinions and confused the story, I tried not to let them influence me. The comments became less important than the story.

I: That's all the questions I have. Is there anything else you'd like to say about the stories or the system?

P: No

I: Thank you for your time.

J.3.2 Participant 6

	Story 1	**Story 2**	**Story 3**
Fair	2	2	2
Unbiased	2	1	2
Wholestory	3	1	2
Accurate	2	2	2
Trust	2	2	2

Table J.6: Participant 6 credibility ratings

This participant believed it was difficult to judge the credibility of the divorce story as it lacked details. They believe that credible stories should tell all relevant points of view, and ensure that no "really clear" facts are missed. They dislike newspapers' tendency to editorialise and to try to make stories more exciting.

They believed the comments followed the story, and tended to pick up the same angles, though in some cases additional information was offered (e.g. for the Formula One story commenters mentioned a video on Youtube). They did see value in reading conversations where people disagree, as it gives insight into more points of view.

They saw and understood the groups, but primarily used the "View All" section to read the comments. They thought that the comments were ordered chronologically and there was no other ordering. They felt that the comments did cover a wide variety of viewpoints but not necessarily complete, and felt that, at least on the divorce story, the comments were more biased than the article.

The participant viewed a fair story as one that represented all sides of an incident while giving enough facts to aid understanding. Their view of bias seemed to be similar, viewing the divorce story as providing insufficient detail to be unbiased. When deciding if they were told the whole story they looked for key parts that were missing. For this participant trust was related to accuracy and to the sources, and accuracy is mostly dependent the source of information: the journalist, the agency, the source of quotations, etc.

I: You read three stories, one about Formula One, one about Divorce, and one about an actress, and you rated them in terms of credibility and gave a summary of the comments and article. What I'm interested in is how you judged the credibility: What you looked at, what made it seem credible, what made it seem non-credible.

P: I didn't get to read everything,

I: It was time limited, that's fine. You can use any of the three stories as an example. First, what did you use when judging the credibility?

P: It's just how I felt about it, how I understood it, and what the comments looked like. For example in the divorce story there aren't many details. They only gave you a little information, you can't really tell what the story is, what's behind it, you can't really judge if it's credible or not.

I: So you think it lacked facts?

P: It lacked facts, yes.

I: Do you mean the story did? or the comments? or both?

P: Both, the comments are based on the story, but sometimes for example in the Formula 1 article, someone said they watched the video on Youtube where it happened. So sometimes the comments also tell you more about the story. So I think in terms of both sometimes you can tell if a journalist is biased or not, if they took just one side it's very clear.

I: So in terms of enhancing credibility, if you could pick out specific things that made you feel something is more credible, what would they be?

P: If the story has two points of views that would make it more credible. I think that's the most important point - and if it didn't miss a really clear fact that is obviously missing.

I: So by that you mean that if you identify something that you think is missing, that would detract from the credibility?

P: Yes, that would detract, exactly.

I: My next question is wha do you think detracted from credibility?

P: It's the opposite. If it doesn't have two sides of the story. And some stories have sentences that make it more attractive, or to incite people. That makes it really non-credible.

I: So from these stories particularly, can you identify anything that you think was missing?

P: For example for the first story [Story 2] it's very shallow, the second one [Story 1] I don't think credibility really matters a lot, and the third one [Story 3] I didn't get to read it carefully enough to be able to judge. Sometimes the comments back up the story.

I: How did the comments change your opinion of the story?

P: People will talk about things that you didn't think of.

I: Do you mean identifying facts that weren't in the story?

P: Yes, identifying facts, identifying the opposite point of view, some people could be extreme. And when people reply to each other that also gives you more.

I: And you found that in all three stories?

P: Mostly

I: You saw the comments in groups.

P: No, I saw the groups. I understand the concept and can tell that they share the same point of view, but I looked at the list at the bottom. I like the colour coding and I could go back and see groups that I wanted to know more about.

I: So did you think there was any meaningful order to the comments?

P: I think they were ordered by the order of the comments. Which is good, because sometimes these are linked. I like it the way it is.

I: And did you feel the comments represented the range of viewpoints that existed.

P: Yes, I would say mostly, but not necessarily completely.

I: Could you identify any that were missing or any that you feel were overrepresented?

P: Not that I can think of, but I would definitely say that they don't completely cover all viewpoints. There's always something missing.

I: So you think in the comments there was a range but not necessarily complete?

P: But there was enough.

I: And would a news system with a larger range be more credible?

P: It would be more credible. Thinking about Twitter for example, when I'm reading about news and politics, you really see different voices appear that you couldn't imagine. So it definitely shows you the different points of view.

I: Finally, I'd like to go through the five points and see how you judged each one. In terms of fairness what would you say is a fair story?

P: A fair story is one which would show you what are the pros and cons or the positives and negatives of the incident or the person. That is what I would call fair and credible. For example with the Lewis Hamilton story I don't really know about the details, so I can't judge, that's why I rated it neutral. With the divorce, it felt a little bit biased and the journalist was trying to incite people. So I don't really know what's happened.

I: So fairness you see as being about showing different sides of an argument or story?

P: Yes, and giving enough facts.

I: And to be unbiased?

P: They don't give enough detail. If you read it and see the comments everyone says "She took his money", so that seems biased.

I: Including the comments?

P: Yes the comments definitely affects that.

I: So were the comments biased in the same way as the article?

P: No, more. The comments were more biased than the article, but the article was slightly.

I: In terms of telling the whole story, what do you look for to tell if an article is telling the whole story?

P: Well, because this was telling a story about divorce, it was just talking about the wealth and the splitting of it - it's a life story, you can never judge people why did she do that, how did she do that. It's just they are trying to make it interesting.

I: So you think there are key parts missing?

P: Yes.

I: Okay, how did you judge if something is accurate?

P: Well accuracy is tricky because I've seen many lies in newspapers, where I was shocked because something that I really know is misrepresented. For example I'd look at who is the journalist, what is the agency, and when you read the news you can tell that some take a particular side. It's very difficult but sometimes you can tell if something is true or not true, for example if they mention where they get information, if they're quoting someone, you can tell sometimes, but it's tricky.

I: So for you it's mostly about the source of the information?

P: Yes

I: Okay, and finally Trust, how do you decide if you should trust a story?

P: I think this is kind of related to accuracy. If it's accurate then it will be trustable, and also the source.

P: Excellent. So what we've been discussing is how you interacted with the stories and in particularly hard you judged if they were credible or not. On that topic is there anything that you think we've missed that's worth saying?

I: Nothing that I can think of. It was fine. The colour coding is brilliant, but maybe it would be confusing the first time you use it.

P: Great. Thank you for your time.

J.4 Condition 4 New Interface New Algorithm

J.4.1 Participant 7

This participant primarily focused on the source of facts in the articles when deciding credibility. They expressed a general doubt about the credibility of news stories, and

	Story 1	**Story 2**	**Story 3**
Fair	1	1	1
Unbiased	1	1	1
Wholestory	2	0	0
Accurate	2	1	1
Trust	1	1	1

Table J.7: Participant 7 credibility ratings

felt that the presence of multiple angles in the story would enhance credibility. They viewed direct quotations as enhancing credibility.

They felt that the angle present in the story influenced the direction of the comments, with the comments primarily focusing on the areas the story did. They did not feel there was a meaningful order to the comments, however they did view the comments as fairly representing the range of views on the subject.

They believed that there was overlap between fairness and being unbiased. When deciding these two areas they were looking for equal representation of all parties and lack of sensationalism. For "whole story" they thought about questions which they had which were not answered in the story. They felt that accuracy depended on the sources of information, and they felt unable to judge trustworthiness without knowing the identity of the original news outlet.

They acknowledged that they were rating purely on the story and not taking the comments into much consideration.

I: When you are evaluating credibility, which elements did you focus on?

P: When they were producing factual information I'm trying to think where that came from and so was there anything to support it? For example in the last one about Sophia Loren the article was saying that it was actually coming from her. She was giving a story about a personal account so I think that might be a little bit more credible if it was coming from her. Though obviously it could still be exaggerated which is why I was a little bit dubious.

I: So you think that the fact that it had direct quotations enhanced the credibility?

P: Yes

I: And what specifically detracted from the credibility? Was it just a general doubt?

P: Maybe I'm just a cynical person when it comes to the news. I'm probably the wrong person to ask

I: What could be added to make it more credible?

P: I think I wrote this as well. With the second story, the worlds most expensive divorce settlement, I did feel it was very one sided. There was a lot of information given about him but the woman was very one dimensional. She was just portrayed as a gold digger, there was nothing abut her and what she's done with her life or anything, so I did feel that detracted from the credibility. Because it didn't give two sides to that story.

I: Did the comments contribute any other angles to that story?

P: I went through quite a bit of the comments, and I got quite angry because there was some which were still questioning him and how he made his money, then there were the anti-feminist comments as well, such as this is why men shouldn't get married. Also she's lived off his money for years so she should only get what she contributed to the marriage. Things like that were just annoying me. I've got pretty strong opinions on things like this, I'm probably going to interpret it and respond in a particular way.

I: So in your opinion, on that story particularly you think both the original story and the comments were biased in the same direction?

P: Well yeah, obviously the story is about him as the subject of the settlement. It's his money. I can see there's going to be a stronger focus on him. But the way it was portrayed I think is going to have an impact on the comments. Like I said there was nothing about her so you're just going to think she's doing well from this marriage and she hasn't done anything herself because there was no information about her.

I: In any of the other two stories, did the comments change your opinion of the original article at all?

P: Not really. I think the comments were quite mixed on all of them. Like with the Formula 1 article you could see there was the backlash between different supporters. I did feel as well that there would be people saying he got away with it. I didn't read the comments and think I'm seeing the story in a different light.

I: Did you think generally the comments fairly represented the range of views on the subject.

P: Yeah it wasn't just an overwhelming majority with one opinion. Apart from the Sophia Loren one where I think most people came back with "oh no she's beautiful". There were different opinions which is interesting and gives different perspectives.

I: Did you feel there was any meaningful order to the comments or was it random?

P: I think it was kind of random. I didn't see a particular order.

I: The last thing I'd like to do is go through the five points and see how you interpreted them: Fair, Unbiased, Tells the Whole Story, Accurate, and Can be Trusted. For fair, what were you looking for to say if something was fair or unfair?

P: I think for me this overlaps with bias really. I'm looking for an equal representation of all parties involved and nothing too sensational. No dramatic wording. You don't need to make it hyperbolic you can just tell the story without trying to make it too emotive.

I: And bias is similar?

P: Probably, like I said trying to make it more impartial. Because I think I'm probably a bit dubious when reading the news, thinking what is the agenda here? If it seems to be more sympathetic to one side rather than the other then I'm going to think that there's more bias involved.

P: So on all three of the stories you said they are quite biased. A second ago you did say that the comments fairly represented the viewpoints. Did you not take that into account in your rating of bias?

I: I think maybe when I was rating it I'm rating it purely on the story and not taking the comments into as much consideration.

P: Okay. So Whole Story how did you interpret that?

I: I read the story and thought "Is everything here?". What are the other questions? and when reading the comments it's nice to see that other people are questioning the same sort of gaps as well. That helped give a more rounded perspective.

P: And Accurate?

I: I feel like there's overlap with fairness and bias. It depends on what the source is. Where the information is from. If they've given valid sources then I'll think the information is more accurate.

P: And finally, Trustworthy

I: For me, this is a news story and I don't know where it's from. Is it from a tabloid or a broadsheet? what political affiliation have they got? I can't help but have my own personal biases so wether or not I'm going to trust it depends on who has written this. I don't know that so that's a difficult one.

J.4.2 Participant 8

	Story 1	**Story 2**	**Story 3**
Fair	4	4	4
Unbiased	3	4	4
Wholestory	3	4	4
Accurate	3	4	4
Trust	4	4	4

Table J.8: Participant 8 credibility ratings

This participant focused heavily on the news outlets. After determining that the three stories were from actual news outlets and not manufactured for the study, the participant placed a lot of trust in the stories. However they did acknowledge that the journalist may have their own biases, and felt that subjectivity would reduce credibility.

They did not feel the comments influenced them much, and saw a strong difference between the "really biased" comments and journalistic news. They did not detect a meaningful order, and felt that too many groups were displayed, overwhelming them with information. They would have preferred a simple "positive comments" and "negative comments" split. However they did feel the comments they read fairly represented the viewpoints.

They were not able to describe how they measure fairness beyond an intuitive feeling, they felt bias related to subjectivity in the article, based their "whole story" and accuracy judgements on prior knowledge of the event, and based trust almost entirely on the news outlet, and on context from other news sources. In this case they did not know the news outlet, and were not provided with context, so they based their trust ratings on how similar the story was to stories they read in outlets they trust.

I: You saw three stories. You then summarised the story and comments, and you rated it in terms of credibility. What I'm interested in now is learning how you approached this, and what you were looking for when you judged credibility. When I ask you these questions feel free to refer to any of the stories as an example. First, what elements did you consider when you were judging credibility?

P: First I looked for the platform. One seemed like Sky News, or a news channel anyway. If it's a news channel it will be credible. One of them wasn't like a news story, it was more like a feature. So I wouldn't know for sure if that piece was written as part of the experiment or if it was an actual news story, because the format and the way it was written was more like a feature story than a news story. So I really couldn't tell if it was news or not. I couldn't

really tell where the news was extracted from, but if I could then I would be able to tell the credibility.

I: So typically you would base it on the platform or the news outlet?

P: The news outlet, yeah. It would be based on the news outlet. That is not to say that all the news outlets would be credible but I would think they have a perception of credibility from a public news outlet.

I: So in the absence of that, what were you using to make your judgements?

P: I was just trying to imagine this story. What was going on in my mind was, this story couldn't have been written for the experiment because it seems like a real story that was extracted, but I couldn't tell from which news outlet. I knew it would be a story, as it would be difficult to concoct such a story so it must be a story that exists in a news outlet.

I: Forgetting about the outlet, because you don't know what the outlet is, was there anything in the stories that you read, or in the comments, or in the appearance or anything else, that you looked at and it made you feel it is credible?

P: Well, with the fact that they are not fictitious character, and Lewis Hamilton is real, Rosberg is real, he's German, Mercedes Benz is German, they are real. This seems to be a real story. But in terms of credibility, I think the story does exist, but the point that I'm looking at is the bias of the writer. Would the writer have been writing from a biased perspective? Especially in this story about Hamilton and Rosberg I could see elements of subjectivity in the writer, not entirely as a news story. In my opinion it was more about opinion, more of a feature story not a news story. If I was going to judge it I wouldn't say it is a credible news story, but it might be a good feature story.

I: Okay, so my next question was is there anything that detracted from credibility. You just answered that as you feel it was subjective, right?

P: Right.

I: Did you feel the divorce story was subjective?

P: The divorce story, I didn't feel it was subjective. That seems a credible story, I didn't sense any subjectivity. One point that I noticed in the comments, the fact that he was known to own Monaco Football Club, probably the suggestion is that he is richer than he declared. So I don't know what that information adds to the story.

I: So how did you affect your judgement?

P: That would make me think probably the writer is more in support of the wife, but maybe there is going to be a subsequent story about Dimitri's assets.

I: Do you think that there is anything missing from these stories, that if they were there, would make you feel it is more credible?

P: Well, the Dimitri story is a bit more straightforward news. It's so difficult to subtract subjectivity completely from any news story, because they are written by humans. The news writer should try as much as possible to just tell it as it is, not what they think.

I: And you mentioned earlier if we told you what the source of the story was, that may have added to credibility?

P: Probably it could. For example Louis Hamilton is English, Rosberg is German. There is this rivalry between them. A German news outlet could tilt towards Rosberg, and an English news outlet would tilt towards Hamilton. So if I knew which outlet this was from I would be able to understand where the balance is tilting.

I: In any cases did the comments change your opinion of the story?

P: Not so much. I didn't get to read much of the comments. Some comments were ones I would have made myself, for example there was one saying they were married for 23 years which means they wouldn't blame the wife for marrying him for his wealth. So she deserves what she got.

I: So you said you didn't read very many of the comments. Of the ones you read did you feel there was any meaningful order or organisation to them?

P: No there is no meaningful order. If I was reading from a news outlet I wouldn't care about the organisation of the comments. I would only care about the comments like if you were reading on Amazon, with 10 positive comments, and 20 negative comments. I would probably want to read the negative comments to make my decision. If the organisation of the comments was in such a way, not too many categories but maybe positive and negative, I would probably understand more.

I: Of the comments you read do you feel they fairly represented the range of viewpoints on the topic?

P: On the divorce story, some were on the side of the men and some were on the side of the woman. There will always be for and against.

I: So do you think it fairly represents the viewpoints?

P: I think it does.

I: And finally, I want to know how you interpreted the five credibility criteria. How did you decide something was fair?

P: The way it's written. I don't know how you measure fairness, I just think it's fair or it's not.

I: We can come back to that. What about bias? how would you judge if a story is biased?

P: From the words the connote the writer's own perception. I would probably think the Lewis Hamilton story was written by an English person because there was more of a tilt towards Louis Hamilton there.

I: Do you think that was biased?

P: Well it could have been biased, towards Hamilton.

I: And for the whole story, how did you judge if an article was telling the whole story?

P: The would be based on previous knowledge of the story. If you have a prior knowledge of the story then you can tell something is not complete. Or if you had access to other news sources. Other that, if you haven't got any prior knowledge you'll be fed what the news story says.

I: How did you just if a story was accurate?

P: Like the previous one this is dependant on your previous understanding of the case, on the amount of information available to you. If you haven't got prior knowledge of the story, and you didn't do further research, you won't be able to judge if it's accurate or not.

I: In these cases did you have prior information?

P: Not at all

I: So you struggled to judge that?

P: Yeah

I: And finally trust, how did you judge if you would trust the system?

P: It would go back to the news outlet, from where I'm reading it, for example an informed mind would know that British media tend to tell it as it is, for example, when it comes to local elections. The BBC for example, goes against the government sometimes, will put out reports against the government. But there are other stories about the Middle East or Asia, where if you really want to trust the credibility you've got to listen to another outlet

like Al-Jazeera. You want to see how Al-Jazeera report the same story, and you also want to see how Russia Today report the same story, because their position will colour the reporting, no matter how much the outlet try to be fair and objective.

I: So in these stories you weren't given a lot of context, you weren't given other stories from other outlets, or told who the outlet is, and so on. In general you've said the three stories are fairly credible and fairly trustworthy. With the absence of context how did you judge?

P: I kind of imagined that these stories couldn't have been made up for the study. I looked at the content of the story. I don't know where it came from but I kind of figured out this would be a real story that existed, and that's why I think it's fairly credible.

I: In those answers you focused primarily on the news outlet and the story, you haven't mentioned the comments at all. Did you take them into account at all or did they not influence you?

P: No they wouldn't influence me because those will be really biased. They are comments. "Comments are free, facts are sacred". There will always be those who are opposing and those who are supporting, and they'll express that through their comments.

Appendix K

Contribution To News Experiment Plan

Though the study in Chapter 6 found no significant impact on the credibility assessments and only limited evidence of an impact on the topics encountered, the effect of the change in interface and algorithm on the way that people interact with and contribute to news has not been determined. This could potentially change the types of people who contribute and the types of contribution made.

Additionally, though the discussions were presented using a novel interface (designed in Chapter 6), the content was produced by people using a traditional news discussion interface. Presenting content produced using novel methods may also lead to different results than those found in Chapter 6.

Future work should investigate the hypothesis that "adding [community-based] structure to open news systems will result in more varied viewpoints contributed". This can be performed by producing a news discussion engine similar in style to Reddit.com, but which presents the discussion using the interface designed in Chapter 6, and then inviting participants to use this system to discuss news stories.

K.1 Experiment Plan

Attempting to run this as a controlled experiment would be very difficult. It would not be possible to test the original and new algorithm on the same group of users without having their experiences of the system with one algorithm influence their perceptions of the system with the other. As the required number of participants and the time commitment required of the participants is high, getting two groups of participants is unlikely to be possible.

Instead, this should be run as an open experiment whereby the modified social news website will be put online for an extended time period, with users able to use the website as they wish throughout that time. There should also be specific events identified and highlighted to encourage participation. After significant contribution has been created, a small number of participants should be randomly selected for follow-up interviews in order to gain an insight into the reasons for the behaviour witnessed.

There are weaknesses with this methodology that must be addressed, the result of decisions taken to make such a study feasible to execute. First, there is the risk that the act of telling the participants that the algorithm has changed will influence behaviour more than the actual change will. This is unavoidable as failing to tell the participants clearly how the algorithm has changed could result in their previous experience and impressions of similar systems overwhelming any effect from the new system and this could stop them engaging with the system in such a way that behavioural change is possible. The participant responses during the study in Chapter 6 indicated that pre-existing views and previous experiences already have a strong influence on future behaviour when consuming news.

There is also the risk that due to the relatively large time commitment required, it will be prohibitively difficult to recruit a large number participants. To reduce the number of required participants as far as possible, a control group will likely be infeasible. Without a control group, only limited conclusions will be able to be drawn from the collected data, though this will allow for comparisons to existing communities and provide insight into possible uses of community detection algorithms in online news communities.

K.2 Design

As mentioned by Springer et al. (2015), there is evidence that features such as registration rules (Tsagkias et al., 2009; Ruiz et al., 2011) and levels of anonymity (Ho and McLeod, 2008) can be influential in users deciding whether to contribute to a news system. To make the evaluation meaningful, it is important that the new interface mirrors existing interfaces as much as possible and only changes that which is necessary for the study.

For implementation, there were two primary options available: use the system created for Chapter 6 and build in additional features to allow participants to contribute to discussions, or use the available source code from Reddit and modify the commenting interface to use the new algorithm and interface. In this case it was decided that using the open-source Reddit source code in this experiment would come with far more features than are required in this study which would complicate deployment. Instead it was decided to add the missing features to the system from Chapter 6.

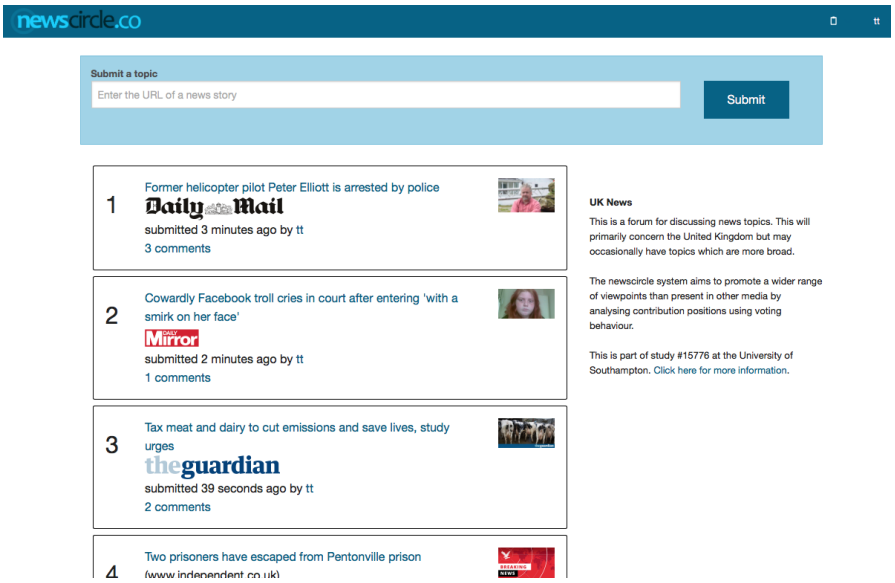


Figure K.1: Index

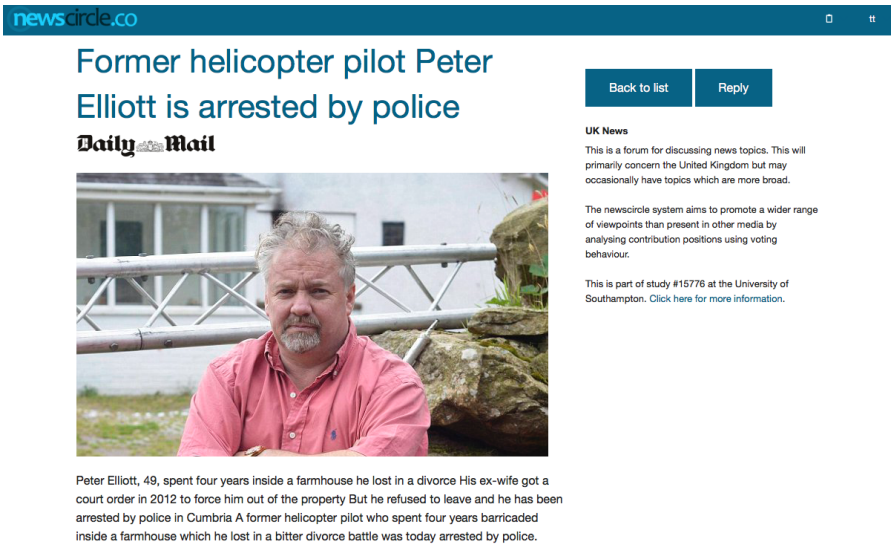


Figure K.2: Topic

The design of the system is heavily inspired by the appearance of Reddit.com and similar systems, making minimal changes to allow for the change of algorithm and to encourage participation. The index page (shown in Figure K.1) lists topics with the most recently active at the top, with instructions on one side and and invitation to submit a topic. For the purposes of this study, the interface has been named "newscircle".

The topic (see Figure K.2) shows the headline of the story as well as a short summary and picture, with a link to read the full story. Below the summary is the discussion interface as used in Chapter 6 (see Figure K.3).

There are two major differences between the interface used in Chapter 6 and the comments displayed on newscircle. First, there is an input at the top of the discussion

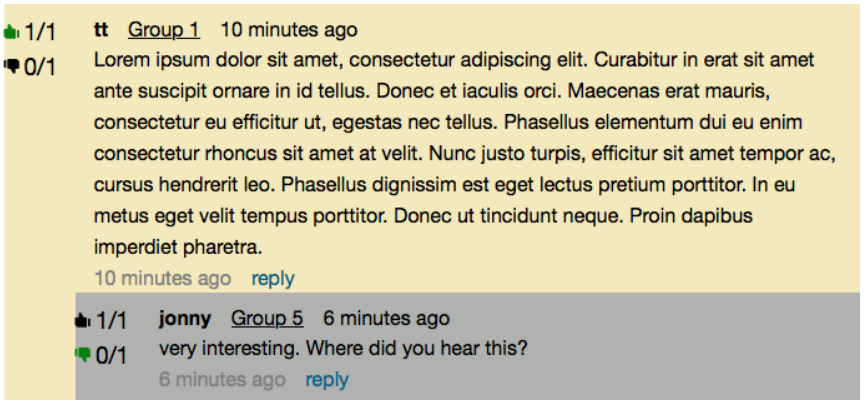


Figure K.3: Comments

Figure K.4: Newscircle voting

area and a toggle beneath each comment inviting further contribution from the readers. Second, to make the interface more similar to existing social news systems the profile pictures have been removed from each contribution and the vote counts moved to the left of the comment. The counts are now shown alongside "thumbs up" and "thumbs down" icons which the reader can click to submit their own vote. This toggle can be seen in Figure K.4.

K.2.1 Ranking Algorithm

On Newscircle, like on Reddit, ranking is performed in two ways: the overall ranking of topics, and the ranking of individual comments within topics.

The overall ranking of topics on Newscircle is much simpler than that of Reddit. On Reddit, story ranking takes into account submission time and number of votes, weighing earlier votes higher than later votes, and penalising controversial topics with a mix of up-votes and down-votes (Salihefendic, 2015). For a Python implementation of Reddit's "hot" algorithm see Figure K.5.

Newscircle does not allow voting on topics and simply orders so that those with more recent comments are shown on top. It does however allow for topics to be hidden, pinned (for instructional topics), or featured (for featured news items to encourage discussion).

The algorithm which ranks comments is the more interesting case for both Reddit and Newscircle. On Reddit, by default comments are ordered by "Best". This uses Wilson Score Interval to estimate the popularity of a comment amongst the entire community

```
from datetime import datetime, timedelta
from math import log

epoch = datetime(1970, 1, 1)

def epoch_seconds(date):
    td = date - epoch
    return td.days * 86400 + td.seconds + (float(td.microseconds) / 1000000)

def score(ups, downs):
    return ups - downs

def hot(ups, downs, date):
    s = score(ups, downs)
    order = log(max(abs(s), 1), 10)
    sign = 1 if s > 0 else -1 if s < 0 else 0
    seconds = epoch_seconds(date) - 1134028003
    return round(sign * order + seconds / 45000, 7)
```

Figure K.5: Reddit Story Ranking Algorithm, simplified by Amir Salihefendic (Salihefendic, 2015)

based on a sample (the votes of people who have voted on it). Reddit's implementation of this can be seen in Figure K.6.

The ranking algorithm used in on NewsCircle runs in two phases. The first phase is the generation of groups. For this it uses the python-louvain[1] implementation of the Louvain algorithm. To ensure that the system can run in real time the groups are regularly re-generated by a separate process and saved in the database. See Figure K.7 for this portion of the algorithm.

The second phase is to order the comments in the same way that they are ordered in the design from Chapter 6. This is with the top comment from each group interleaved so that each is represented early on in the discussion. This algorithm uses a simple method of deciding the top comment in each group, by subtracting dislikes from likes. The full source code for this implementation is available at `https://github.com/jscott1989/newscircle`.

K.2.2 Changed Functionality

To make it as easy as possible for participants to begin new topics, they are first invited to enter the URL of a story (see Figure K.8) and then the title, image, and description is pre-populated from the URL (see Figure K.9). This makes it as easy as possible for

[1] Available from `http://perso.crans.org/aynaud/communities/`

```
from math import sqrt

def _confidence(ups, downs):
    n = ups + downs

    if n == 0:
        return 0

    z = 1.281551565545
    p = float(ups) / n

    left = p + 1/(2*n)*z*z
    right = z*sqrt(p*(1-p)/n + z*z/(4*n*n))
    under = 1+1/n*z*z

    return (left - right) / under

def confidence(ups, downs):
    if ups + downs == 0:
        return 0
    else:
        return _confidence(ups, downs)
```

Figure K.6: Reddit Comment Ranking Algorithm, simplified by Amir Salihefendic (Salihefendic, 2015)

participants to create new topics but does not allow for the creation of topics without a story (as Reddit does). For the purposes of this study this will suffice as more direct comparisons can be made between discussions about the same story on different systems.

When creating accounts, users are able to log in instantly using their Google, Facebook, or Twitter accounts, or otherwise provide a username and password. To encourage participation, the "New Topic" and "Reply" functionality is enabled even for users who are not logged in, however before submitting the content they will be prompted to create an account. For an example see Figure K.10.

This means that readers are offered the entire functionality of Newscircle without needing to register, and registration is required at the latest possible time. This should ensure maximum participation.

K.3 Data Analysis

The perceived credibility of the system can be compared to Reddit or a similar online news community using the methods employed in Chapter 4 and Chapter 6. The study in Chapter 6 found the new algorithm or interface resulted in no difference to credibility

```
users = {}
for comment in topic.comments.all():
    # Everyone who liked this plus the author get put into a set
    likers = set([
        liker.id for liker in comment.liked_by.all()] +
        [comment.author.id])

    # Dislikers in another set
    dislikers = set([disliker.id for disliker in comment.disliked_by.all()])

    # Users who like the same thing get increased relationship
    for usera, userb in [b for b in
            itertools.permutations(likers, 2) if b[0] < b[1]]:
        change_relationship(users, usera, userb, 1)

    # Users who dislike the same thing get increased relationship
    for usera, userb in [b for b in
            itertools.permutations(dislikers, 2) if b[0] < b[1]]:
        change_relationship(users, usera, userb, 1)

    # Users who like/dislike the opposite thing get decreased relationship
    for usera, userb in itertools.product(likers, dislikers):
        change_relationship(users, usera, userb, -1)

g = Graph()

for user, relationships in users.items():
    g.add_node(user)
    for weight in relationships.values():
        g.add_edge(user, r, weight=weight)

communities = {}
partition = community_finder.best_partition(g)
for user_id, community_id in partition.items():
    if community_id not in communities:
        communities[community_id] = []
    communities[community_id].append(user_id)

communities = communities[:7]

# Save communities
```

Figure K.7: Newscircle Comment Ranking Algorithm

Figure K.8: URL

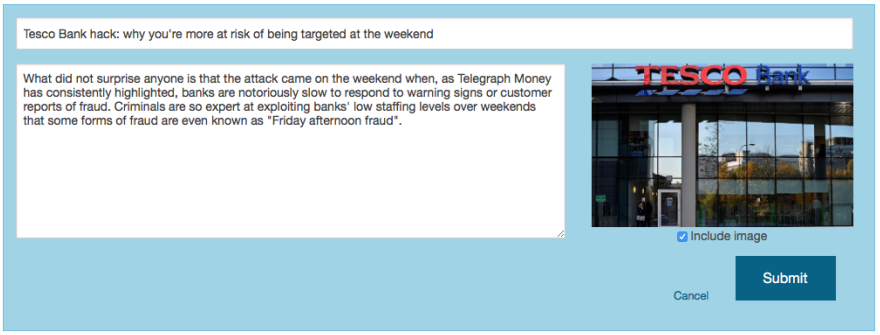


Figure K.9: Article

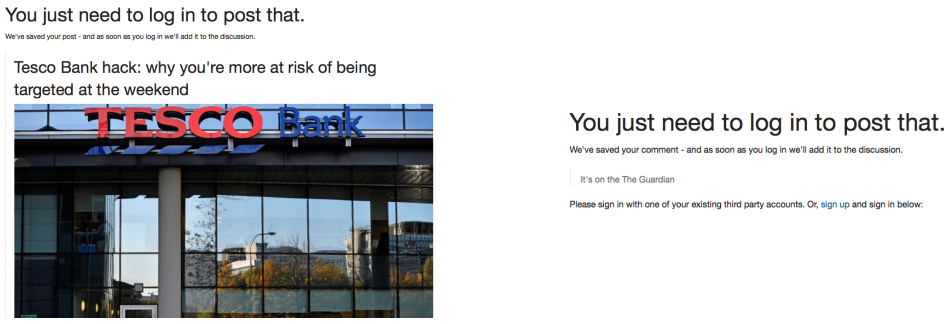


Figure K.10: Newscircle signup

however the result may be different when the content is produced using the new system rather than being repurposed from existing systems. Having participants who were engaged in the discussion rate the credibility may also ensure that they have a more complete understanding of how the system works.

Additionally, discussions can be identified and equivalent topics collected from Reddit. The structure of these discussions can then be directly compared. This will involve comparing the contributions per person, the sizes of the communities within the discussions, the average number of responses and depth of comments, and other metrics to indicate if the structure of the discussion has changed due to the change in interface and algorithm. Reddit has a very large number of subreddits[2] and the proposed system should support a small subset of equivalent categories so that direct comparisons can be made.

K.4 Semi-structured interview

While the content will be analysed with an interest in properties of the content produced, semi-structured interviews will be used to gain qualitative information about the ways that people interacted with the system. The questions are based on the four dimensions

[2]Subreddit is the name given to a category on Reddit. As of June 2015 Reddit had aproximately 10,000 active subreddits according to The Reddit blog at `https://redditblog.com/2015/06/23/happy-10th-birthday-to-us-celebrating-the-best-of-10-years-of-reddit/`

of motivations as detailed in Springer et al. (2015): Cognitive, Entertainment, Social-Integrative, and Personal Identity, and some demographic questions to give context to the answers. As this is a semi-structured interview, the answers given to the questions will allow for further questions to investigate differences between the system used and existing systems participants have experience with.

Demographic

1. What online systems do you typically use for discussing news topics?
2. How often did you use the discussion system in this study?
3. How often did you vote?
4. How often did you comment?

Cognitive

1. Do you feel you were able to share your experiences with other people using the system?
2. Do you feel you were given an opportunity to provide corrections or balance to the article?
3. Do you feel you could contribute additional information about the topic?

Entertainment

1. Do you feel the discussion on the system tended to be aggressive, insulting, or offensive?
2. Were you ever aggressive, insulting, or offensive?

Social-Integrative

1. Did you feel you could interact and socialise with others?
2. Did you feel inclusion or belonging to the community?
3. What is your understanding of how the groups were formed?
4. What did you think of the groups you were placed in?
5. Did you attempt to manipulate the group you were placed in?

6. How much confidence did you have in expressing a contrary view?

Personal Identity

1. Were your own opinions fairly represented within the community?
2. Were you given the opportunity to express your opinion?
3. Were your contributions acknowledged?
4. How well do you think you could identify the attitudes of others?
5. Do you feel you changed anyone else's views through your interaction?

This totals 20 questions. These will be used to follow up with 10-15 participants after they have interacted with the system in order to explain why they behaved in the way that they did. The answers can be compared to the observed behaviours of the participants.

References

Ali, S. R. and Fahmy, S. (2013). Gatekeeping and citizen journalism: The use of social media during the recent uprisings in iran, egypt, and libya. *Media, War & Conflict*, 6(1):55–69.

An, J., Cha, M., Gummadi, K., and Crowcroft, J. (2011). Media landscape in twitter: A world of new conventions and political diversity. In *Proceedings of the Fifth International Conference on Weblogs and Social Media*, pages 18–25, Menlo Park, CA. AAAI.

An, J., Quercia, D., Cha, M., Gummadi, K., and Crowcroft, J. (2014). Sharing political news: the balancing act of intimacy and socialization in selective exposure. *EPJ Data Science*, 3(1):1–21.

Anderson, P. J., Williams, M., and Ogola, G. (2013). *The Future of Quality News Journalism: A Cross-Continental Analysis*. Routledge, New York, NY, 10001.

Arceneaux, N. and Weiss, A. S. (2010). Seems stupid until you try it: press coverage of twitter, 2006-9. *New Media & Society*, 12(8):1262–1279.

Arguello, J., Butler, B. S., Joyce, E., Kraut, R., Ling, K. S., Rosé, C., and Wang, X. (2006). Talk to me: Foundations for successful individual-group interactions in online communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 959–968, New York, NY, USA. ACM.

Armstrong, C. L. and Collins, S. J. (2009). Reaching out: Newspaper credibility among young adult readers. *Mass Communication and Society*, 12(1):97–114.

Arnstein, S. R. (1969). A ladder of citizen participation. *Journal of the American Institute of planners*, 35(4):216–224.

Bardoel, J. (1996). Beyond journalism: A profession between information society and civil society. *European Journal of Communication*, 11(3):283–302.

Becker, H. S. (1967). Whose side are we on? *Social Problems*, 14(3):239–247.

Beckett, C. (2010). The value of networked journalism. Lse research online documents on economics, London School of Economics and Political Science, LSE Library.

Benkler, Y. (2011). Free irresponsible press: Wikileaks and the battle over the soul of the networked fourth estate. *Harvard Civil Rights-Civil Liberties Law Review*, 46:311.

Bennett, W. L. (1988). *News: The Politics of Illusion*. Longman, New York.

Bentley, C. (2001). No newspaper is no fun–even five decades later. *Newspaper Research Journal*, 22(4):2–15.

Berelson, B. (1948). What 'missing the newspaper' means. *Communications research 1948-1949*.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10).

Boorstin, D. J. (1961). *The image: A guide to pseudo-events in America*. Random House LLC, New York, United States.

Bowman, S. and Willis, C. (2003). We media: How audiences are shaping the future of news and information. *The Media Center at the American Press Institute*.

Bradshaw, P. (2007). Wiki journalism: Are wikis the new blogs. In *Future of Newspaper Conference*, Cardiff, UK.

Breed, W. (1955). Social control in the newsroom: A functional analysis. *Social forces*, pages 326–335.

Broersma, M. and Graham, T. (2013). Twitter as a news source: How dutch and british newspapers used tweets in their news coverage, 2007–2011. *Journalism Practice*, 7(4):446–464.

Brooke, J. (1996). Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7.

Bruns, A. (2005). *Gatewatching: Collaborative online news production*. Peter Lang, New York.

Bruns, A. (2006). Wikinews: The next generation of online news? *Scan Journal*, 3(1).

Bruns, A. (2007). Produsage. In *Proceedings of the 6th ACM SIGCHI conference on Creativity & cognition*, pages 99–106. ACM.

Bruns, A. (2010). News produsage in a pro-am mediasphere: Why citizen journalism matters. pages 132–147. Palgrave Macmillan, London.

Bruns, A. (2015). Working the story: News curation in social media as a second save of citizen journalism. *The Routledge Companion to Alternative and Community Media*, pages 379–388.

Bruns, A. and Highfield, T. (2012). Blogs, twitter, and breaking news : the produsage of citizen journalism. In Lind, R. A., editor, *Produsing Theory in a Digital World : The Intersection of Audiences and Production in Contemporary Theory*, Digital Formations, pages 15–32. Peter Lang Publishing Inc., New York.

Carter, R. F. and Greenberg, B. S. (1965). Newspapers or television: which do you believe? *Journalism & Mass Communication Quarterly*, 42(1):29–34.

Cassidy, W. P. (2007). Online news credibility: An examination of the perceptions of newspaper journalists. *Journal of Computer-Mediated Communication*, 12(2):478–498.

Champlin, D. P. and Knoedler, J. T. (2006). The media, the news, and democracy: Revisiting the dewey-lippman debate. *Journal of Economic Issues*, pages 135–152.

Chang, L. K. and Lemert, J. B. (1968). The invisible newsman and other factors in media competition. *Journalism & Mass Communication Quarterly*, 45(3):436–444.

Cheng, J., Danescu-Niculescu-Mizil, C., and Leskovec, J. (2014). How community feedback shapes user behavior. *ICWSM 2014*.

Choi, J. H., Watt, J. H., and Lynch, M. (2006). Perceptions of news credibility about the war in iraq: Why war opponents perceived the internet as the most credible medium. *Journal of Computer-Mediated Communication*, 12(1):209–229.

Chung, C. J., Nam, Y., and Stefanone, M. A. (2012). Exploring online news credibility: The relative influence of traditional and technological factors. *Journal of Computer-Mediated Communication*, 17(2):171–186.

Chung, D. S. (2008). Interactive features of online newspapers: Identifying patterns and predicting use of engaged readers. *Journal of Computer-Mediated Communication*, 13(3):658–679.

Clarke, C. L. A., Kolla, M., Cormack, G. V., Vechtomova, O., Ashkan, A., Büttcher, S., and MacKinnon, I. (2008). Novelty and diversity in information retrieval evaluation. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '08, pages 659–666, New York, NY, USA. ACM.

Clauset, A., Newman, M. E., and Moore, C. (2004). Finding community structure in very large networks. *Physical review E*, 70(6):066111.

Cook, E., Teasley, S. D., and Ackerman, M. S. (2009). Contribution, commercialization & audience: understanding participation in an online creative community. In *Proceedings of the ACM 2009 international conference on Supporting group work*, pages 41–50. ACM.

Cook, J. E. and Attari, S. Z. (2012). Paying for what was free: Lessons from the new york times paywall. *Cyberpsychology, Behavior, and Social Networking*, 15(12):682–687.

Crowley, D. N., Breslin, J. G., Corcoran, P., and Young, K. (2012). Gamification of citizen sensing through mobile social reporting. In *IEEE International Games Innovation Conference*, pages 1–5. IEEE.

Curran, J. and Seaton, J. (2009). *Power without responsibility: press, broadcasting and the internet in Britain*. Routledge.

Daniel, A. and Flew, T. (2010). The guardian reportage of the uk mp expenses scandal: A case study of computational journalism. In *Record of the Communications Policy and Research Forum 2010*, pages 186–194. Network Insight Pty. Ltd.

Dave, K., Wattenberg, M., and Muller, M. (2004). Flash forums and forumreader: navigating a new kind of large-scale online discussion. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 232–241. ACM.

Davies, N. (2011). *Flat Earth news: an award-winning reporter exposes falsehood, distortion and propaganda in the global media*. Random House, London.

Deuze, M. (2008). The professional identity of journalists in the context of convergence culture. *Observatorio (Obs*)*, 2(4).

Diakopoulos, N., De Choudhury, M., and Naaman, M. (2012). Finding and assessing social media information sources in the context of journalism. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2451–2460. ACM.

Diakopoulos, N. and Koliska, M. (2016). Algorithmic transparency in the news media. *Digital Journalism*, pages 1–20.

Diakopoulos, N. and Naaman, M. (2011). Towards quality discourse in online news comments. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, CSCW '11, pages 133–142, New York, NY, USA. ACM.

Domingo, D., Quandt, T., Heinonen, A., Paulussen, S., Singer, J. B., and Vujnovic, M. (2008). Participatory journalism practices in the media and beyond: An international comparative study of initiatives in online newspapers. *Journalism practice*, 2(3):326–342.

Downie, L. and Schudson, M. (2009). The reconstruction of american journalism. *Columbia Journalism Review*, 19.

Eason, D. L. (1986). On journalistic authority: The janet cooke scandal. *Critical Studies in Mass Communication*, 3(4):429–447.

Elmer, G. (2012). Live research: Twittering an election debate. *New Media & Society*.

Evans, H. (1972). *Editing and Design: Newsman's English*, volume 1. Butterworth-Heinemann Ltd, London: Heinemann.

Faridani, S., Bitton, E., Ryokai, K., and Goldberg, K. (2010). Opinion space: A scalable tool for browsing online comments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 1175–1184, New York, NY, USA. ACM.

Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4):82–89.

Fenton, N. (2010). *New media, old news: Journalism and democracy in the digital age*. Sage Publications.

Flanagin, A. J. and Metzger, M. J. (2000). Perceptions of internet information credibility. *Journalism & Mass Communication Quarterly*, 77(3):515–540.

Flanagin, A. J. and Metzger, M. J. (2007). The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. *New Media & Society*, 9(2):319–342.

Flavián, C. and Gurrea, R. (2006). The choice of digital newspapers: influence of reader goals and user experience. *Internet Research*, 16(3):231–247.

Fogg, B. (2003). Prominence-interpretation theory: Explaining how people assess credibility online. In *CHI'03 extended abstracts on human factors in computing systems*, pages 722–723. ACM.

Fogg, B. J. and Tseng, H. (1999). The elements of computer credibility. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, pages 80–87, New York, NY, USA. ACM.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(3):75–174.

Fortunato, S. and Barthelemy, M. (2007). Resolution limit in community detection. *Proceedings of the National Academy of Sciences*, 104(1):36–41.

Fowler, R. (1991). *Language in the News*. Routledge, London.

France, M. (1999). Journalism's online credibility gap. *Business Week*, 3650:122–124.

Freelon, D. (2015). Discourse architecture, ideology, and democratic norms in online political discussion. *New Media & Society*, 17(5):772–791.

Frey, D. (1986). Recent research on selective exposure to information. *Advances in experimental social psychology*, 19(1):41–80.

Frost, C. (2015). *Journalism ethics and regulation*. Routledge.

Galtung, J. and Ruge, M. H. (1965). The structure of foreign news the presentation of the congo, cuba and cyprus crises in four norwegian newspapers. *Journal of peace research*, 2(1):64–90.

Gans, H. J. (1979). *Deciding what's news: A study of CBS evening news, NBC nightly news, Newsweek, and Time*. Northwestern University Press.

Gans, H. J. (2011). Multiperspectival news revisited: Journalism and representative democracy. *Journalism*, 12(1):3–13.

Garrett, R. K. (2009). Echo chambers online?: Politically motivated selective exposure among internet news users1. *Journal of Computer-Mediated Communication*, 14(2):265–285.

Gaziano, C. and McGrath, K. (1986). Measuring the concept of credibility. *Journalism Quarterly*, 63(3):451–462.

Gentzkow, M., Glaeser, E. L., and Goldin, C. (2006). The rise of the fourth estate. how newspapers became informative and why it mattered. In *Corruption and Reform: Lessons from America's Economic History*, pages 187–230. University of Chicago Press.

Gentzkow, M. and Shapiro, J. M. (2010). Ideological segregation online and offline. Technical report, National Bureau of Economic Research.

Giannopoulos, G., Koniaris, M., Weber, I., Jaimes, A., and Sellis, T. (2014). Algorithms and criteria for diversification of news article comments. *Journal of Intelligent Information Systems*, 44(1):1–47.

Girvan, M. and Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826.

Greenwald, G. (2014). *No place to hide: Edward Snowden, the NSA, and the US surveillance state*. Macmillan.

Groseclose, T. and Milyo, J. (2005). A measure of media bias. *The Quarterly Journal of Economics*, pages 1191–1237.

Haas, T. (2005). From "public journalism" to the "public's journalism"? rhetoric and reality in the discourse on weblogs. *Journalism Studies*, 6(3):387–396.

Hall, S., Clarke, J., Critcher, C., Jefferson, T., and Roberts, B. (1978). *Policing the crisis: Mugging, law and order and the state*. London: Macmillan.

Hampton, K., Goulet, L. S., Rainie, L., and Purcell, K. (2011). Social networking sites and our lives. Technical report, Pew Internet and American Life Project.

Hänska-Ahy, M. T. and Shapour, R. (2013). Who's reporting the protests? converging practices of citizen journalists and two bbc world service newsrooms, from iran's election protests to the arab uprisings. *Journalism Studies*, 14(1):29–45.

Harcup, T. and O'Neill, D. (2001). What is news? galtung and ruge revisited. *Journalism Studies*, 2(2):261–280.

Hecht, B. J. and Gergle, D. (2010). On the localness of user-generated content. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pages 229–232. ACM.

Hellmueller, L. and Trilling, D. (2012). The credibility of credibility measures: A meta-analysis in leading communication journals, 1951 to 2011. In *WAPOR 65th Annual Conference in Hong Kong*.

Heravi, B. R., Boran, M., and Breslin, J. G. (2012). Towards social semantic journalism. In *Workshop on the Potential of Social Media Tools and Data for Journalism in News and Media Industry at the Sixth International AAAI Conference on Weblogs and Social Media*.

Herman, E. S. and Chomsky, N. (1988). *Manufacturing consent: The political economy of the mass media*. Random House, London.

Hermida, A., Domingo, D., Heinonen, A., Paulussen, S., Quandt, T., Reich, Z., Singer, J. B., and Vujnovic, M. (2011). The active recipient: Participatory journalism through the lens of the dewey-lippmann debate. In *International Symposium on Online Journalism*, volume 1.

Hermida, A., Lewis, S. C., and Zamith, R. (2014). Sourcing the arab spring: A case study of andy carvin's sources on twitter during the tunisian and egyptian revolutions. *Journal of Computer-Mediated Communication*, 19(3):479–499.

Hermida, A. and Thurman, N. (2007). Comments please: How the british news media are struggling with user-generated content. In *Articulo presentado en el 8th International Symposium on Online Journalism*.

Hilligoss, B. and Rieh, S. Y. (2008). Developing a unifying framework of credibility assessment: Construct, heuristics, and interaction in context. *Information Processing & Management*, 44(4):1467–1484.

Ho, S. S. and McLeod, D. M. (2008). Social-psychological influences on opinion expression in face-to-face and computer-mediated communication. *Communication Research*, 35(2):190–207.

Holmqvist, K., Holsanova, J., Barthelson, M., and Lundqvist, D. (2003). *Reading or scanning? A study of newspaper and net paper reading.*, pages 657–670. The mind's eye: cognitive and applied aspects of eye movement research. Elsevier Science Ltd.

Hoque, E. and Carenini, G. (2015). Convisit: Interactive topic modeling for exploring asynchronous online conversations. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*, pages 169–180. ACM.

Horrigan, J. B., Garrett, K., and Resnick, P. (2004). *The Internet and democratic debate*. Pew Internet & American Life Project.

Iivari, N. (2004). Enculturation of user involvement in software development organizations - an interpretive case study in the product development context. In *Proceedings of the Third Nordic Conference on Human-computer Interaction*, NordiCHI '04, pages 287–296, New York, NY, USA. ACM.

Ipsos MORI (2013). Perceptions are not reality: The top 10 we get wrong. Accessed at `https://www.ipsos-mori.com/researchpublications/researcharchive/3188/Perceptions-are-not-reality-the-top-10-we-get-wrong.aspx`.

Iyengar, S. and Hahn, K. S. (2009). Red media, blue media: Evidence of ideological selectivity in media use. *Journal of Communication*, 59(1):19–39.

Jaffali, S., Jamoussi, S., and Hamadou, A. B. (2014). Grouping like-minded users based on text and sentiment analysis. In *Computational Collective Intelligence. Technologies and Applications*, pages 83–93. Springer.

Jarman, R. and McClune, B. (2007). *Developing Scientific Literacy: Using News Media In The Classroom*. McGraw-Hill International.

Johnson, M. and Hyysalo, S. (2012). Lessons for participatory designers of social media: long-term user involvement strategies in industry. In *Proceedings of the 12th Participatory Design Conference: Research Papers-Volume 1*, pages 71–80. ACM.

Johnson, T. J. and Kaye, B. K. (1998). Cruising is believing?: Comparing internet and traditional sources on media credibility measures. *Journalism & Mass Communication Quarterly*, 75(2):325–340.

Johnson, T. J. and Kaye, B. K. (2004). Wag the blog: How reliance on traditional media and the internet influence credibility perceptions of weblogs among blog users. *Journalism & Mass Communication Quarterly*, 81(3):622–642.

Johnson, T. J. and Kaye, B. K. (2009). In blog we trust? deciphering credibility of components of the internet among politically interested internet users. *Computers in Human Behavior*, 25(1):175–182.

Johnson, T. J., Kaye, B. K., Bichard, S. L., and Wong, W. J. (2007). Every blog has its day: Politically-interested internet users' perceptions of blog credibility. *Journal of Computer-Mediated Communication*, 13(1):100–122.

Jones, D. A. (2004). Why americans don't trust the media: A preliminary analysis. *The Harvard International Journal of Press/Politics*, 9(2):60–75.

Karlsson, M., Clerwall, C., and Nord, L. (2014). You ain't seen nothing yet. *Journalism Studies*, 15(5):668–678.

Keegan, B. and Gergle, D. (2010). Egalitarians at the gate: One-sided gatekeeping practices in social media. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pages 131–134. ACM.

Kim, D. and Johnson, T. J. (2009). A shift in media credibility comparing internet and traditional news sources in south korea. *International Communication Gazette*, 71(4):283–302.

King, G., Pan, J., and Roberts, M. E. (2013). How censorship in china allows government criticism but silences collective expression. *American Political Science Review*, 107(02):326–343.

Kiousis, S. (2001). Public trust or mistrust? perceptions of media credibility in the information age. *Mass Communication & Society*, 4(4):381–403.

Kobayashi, T. and Ikeda, K. (2009). Selective exposure in political web browsing: Empirical verification of 'cyber-balkanization'in japan and the usa. *Information, Communication & Society*, 12(6):929–953.

Kohring, M. and Matthes, J. (2007). Trust in news media: Development and validation of a multidimensional scale. *Communication Research*, 34(2):231–252.

Kohut, A. (2011). Views of the news media: 1985-2011. *Pew Research Center*. Accessed at `http://www.people-press.org/2011/09/22/press-widely-criticized-but-trusted-more-than-other-institutions/`.

Kohut, A., Doherty, D. C., Dimock, M., and Keeter, S. (2008). Key news audiences now blend online and traditional sources. *Pew Research Center Biennial News Consumption Survey*, 1:2009.

Kovach, B. and Rosenstiel, T. (2007). *The elements of journalism: What newspeople should know and the public should expect*. Random House LLC.

Kovačič, M. P. and Erjavec, K. (2008). Mobi journalism in slovenia: Is this really citizen journalism? *Journalism Studies*, 9(6):874–890.

Kperogi, F. A. (2011). Cooperation with the corporation? cnn and the hegemonic cooptation of citizen journalism through ireport. com. *New Media & Society*, 13(2):314–329.

Kriplean, T., Toomim, M., Morgan, J., Borning, A., and Ko, A. (2012). Is this what you meant?: promoting listening on the web with reflect. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1559–1568. ACM.

Kwak, H., Lee, C., Park, H., and Moon, S. (2010). What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, pages 591–600. ACM.

Lasica, J. D. (2003). What is participatory journalism. *Online Journalism Review*, 7(08):2003.

Lee, E.-J. and Shin, S. Y. (2012). Are they talking to me? cognitive and affective effects of interactivity in politicians' twitter communication. *Cyberpsychology, Behavior, and Social Networking*, 15(10):515–520.

Lerman, K. and Ghosh, R. (2010). Information contagion: An empirical study of the spread of news on digg and twitter social networks. *ICWSM*, 10:90–97.

Leung, L. (2009). User-generated content on the internet: an examination of gratifications, civic engagement and psychological empowerment. *New Media & Society*, 11(8):1327–1347.

Lewis, C. and Rieman, J. (1993). *Task-centered user interface design: A Practical Introduction.*

Lewis, S. C. (2012). The tension between professional control and open participation: Journalism and its boundaries. *Information, Communication & Society*, 15(6):836–866.

Lewis, T. (1992). "a godlike presence": The impact of radio on the 1920s and 1930s. *OAH Magazine of History*, 6(4):26–33.

Lih, A. (2004). Wikipedia as participatory journalism: Reliable sources? metrics for evaluating collaborative media as a news resource. *Nature.*

Lippmann, W. (1920). *Liberty and the News.* Courier Corporation.

Lippmann, W. (1922). *Public opinion.* Harcourt, Brace and Co., New York.

Loosen, W. and Schmidt, J.-H. (2012). (re-) discovering the audience: The relationship between journalism and audience in networked digital media. *Information, Communication & Society*, 15(6):867–887.

Lord Justice Leveson (2012). *An Inquiry Into the Culture, Practices and Ethics of the Press: Executive Summary and Recommendations [Leveson Report].* The Stationery Office, London.

Lüders, M., Prøitz, L., and Rasmussen, T. (2010). Emerging personal media genres. *New Media & Society.*

Ludford, P. J., Cosley, D., Frankowski, D., and Terveen, L. (2004). Think different: Increasing online community participation using uniqueness and group dissimilarity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 631–638, New York, NY, USA. ACM.

Lukes, S. (1974). *Power: A radical view.* London Macmillan.

MacDougall, C. D. and Reid, R. D. (1968). *Interpretative Reporting.* Macmillan, New York.

Malin, B. J. (2011). A very popular blog: The internet and the possibilities of publicity. *New Media & Society*, 13(2):187–202.

Marchionni, D. M. (2013). Measuring conversational journalism: An experimental test of wiki, twittered and "collaborative" news models. *Studies in Media and Communication*, 1(2):119–131.

Marshall, C. C. (2007). The gray lady gets a new dress: a field study of the times news reader. In *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*, pages 259–268. ACM.

Matthiesson, S. (2010). Who's afraid of the limelight? the trafigura and terry super-injunctions, and the subsequent fallout. *Journal of Media Law*, 2(2):153–167.

McCombs, M. (2005). A look at agenda-setting: Past, present and future. *Journalism studies*, 6(4):543–557.

McCombs, M. (2013). *Setting the agenda: The mass media and public opinion*. John Wiley & Sons.

McCombs, M. E. and Shaw, D. L. (1972). The agenda-setting function of mass media. *Public opinion quarterly*, 36(2):176–187.

McCroskey, J. C. and Young, T. J. (1979). The use and abuse of factor analysis in communication research. *Human Communication Research*, 5(4):375–382.

McQuail, D. (1992). *Media performance: Mass communication and the public interest*. Sage.

Meier, F., Elsweiler, D., and Wilson, M. L. (2014). More than liking and bookmarking? towards understanding twitter favouriting behaviour. *ICWSM 2014*.

Melican, D. B. and Dixon, T. L. (2008). News on the net credibility, selective exposure, and racial prejudice. *Communication Research*, 35(2):151–168.

Metzger, M. J. (2007). Making sense of credibility on the web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology*, 58(13):2078–2091.

Metzger, M. J., Flanagin, A. J., Eyal, K., Lemus, D. R., and McCann, R. M. (2003). Credibility for the 21st century: Integrating perspectives on source, message, and media credibility in the contemporary media environment. *Communication yearbook*, 27:293–336.

Metzger, M. J., Flanagin, A. J., and Medders, R. B. (2010). Social and heuristic approaches to credibility evaluation online. *Journal of Communication*, 60(3):413–439.

Meyer, P. (1988). Defining and measuring credibility of newspapers: Developing an index. *Journalism & Mass Communication Quarterly*, 65(3):567–574.

Mills, R. (2011). Researching social news – is reddit.com a mouthpiece for the 'hive mind', or a collective intelligence approach to information overload? In *ETHICOMP 2011 Proceedings*. Sheffield Hallam University.

Mitchell, A., Gottfried, J., Barthel, M., and Shearer, E. (2016). The modern news consumer. *Pew Research Center*.

Mitchelstein, E. and Boczkowski, P. J. (2010). Online news consumption research: An assessment of past work and an agenda for the future. *New Media & Society*.

Muchnik, L., Aral, S., and Taylor, S. J. (2013). Social influence bias: A randomized experiment. *Science*, 341(6146):647–651.

Muller, M. J. (2003). Participatory design: the third space in hci. *Human-computer interaction: Development process*, pages 165–185.

Munson, S. A. and Resnick, P. (2010). Presenting diverse political opinions: how and how much. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1457–1466. ACM.

Murdock, G. (1974). Mass communication and the construction of meaning. *Reconstructing social psychology*, pages 205–220.

Mustafaraj, E., Metaxas, P. T., Finn, S., and Monroy-Hernández, A. (2012). Hiding in plain sight: A tale of trust and mistrust inside a community of citizen reporters. In *ICWSM*.

Mutz, D. C. (2001). Facilitating communication across lines of political difference: The role of mass media. *American Political Science Association*, pages 97–114.

Nam, K. K. and Ackerman, M. S. (2007). Arkose: reusing informal information from online discussions. In *Proceedings of the 2007 international ACM conference on Supporting group work*, pages 137–146. ACM.

Nardelli, A. and Arnett, G. (2014). You are probably wrong about almost everything. *The Guardian*, 29 October 2014. Accessed at `http://www.theguardian.com/news/datablog/2014/oct/29/todays-key-fact-you-are-probably-wrong-about-almost-everything`.

Nemeth, C. (1986). Differential contributions of majority and minority influence. *Psychological Review*, 93(1):23–32.

Nemeth, C. and Rogers, J. (1996). Dissent and the search for information. *British Journal of Social Psychology*, 35(1):67–76.

Newman, M. E. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113.

Newman, N. (2009). The rise of social media and its impact on mainstream journalism. *Reuters Institute for the Study of Journalism*, 8(2):1–5.

Newman, N., Fletcher, R., Levy, D. A., and Nielsen, R. K. (2016). Reuters institute digital news report 2016. *Available at SSRN: http://ssrn.com/abstract=2796534*.

Nip, J. Y. (2006). Exploring the second phase of public journalism 1. *Journalism Studies*, 7(2):212–236.

Nordenstreng, K. (1971). *A Policy of News Transmission*. National Association of Educational Broadcasters.

Norris, P. (2001). *Digital divide: Civic engagement, information poverty, and the Internet worldwide*. Cambridge University Press.

Obijiofor, L. (2015). African journalists as an endangered breed. *African Journalism Studies*, 36(1):122–128.

Ofcom (2015). News consumption in the uk. Technical report, Ofcom.

Olaisen, J. (1990). Information quality factors and the cognitive authority of electronic information. In *Information Quality: Definitions and Dimensions*, pages 99–120. Taylor Graham, Los Angeles.

Olson, K. W. (2003). *Watergate: the presidential scandal that shook America*. University Press of Kansas.

O'Reilly, T. (2005). What is web 2.0? design patterns and business models for the next generation of software.

Papandrea, M.-R. (2006). Citizen journalism and the reporter's privilege. *Minnesota Law Review*, 91:515.

Parau, P., Stef, A., Lemnaru, C., Dinsoreanu, M., and Potolea, R. (2013). Using community detection for sentiment analysis. In *2013 IEEE International Conference on Intelligent Computer Communication and Processing*, pages 51–54. IEEE.

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin UK.

Park, D., Sachar, S., Diakopoulos, N., and Elmqvist, N. (2016). Supporting comment moderators in identifying high quality online news comments. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 1114–1125. ACM.

Paulussen, S. and Ugille, P. (2008). User generated content in the newsroom: Professional and organisational constraints on participatory journalism. *Westminster Papers in Communication and Culture*, 5(2):24–41.

Pavlik, J. V. (2013). Trends in new media research: A critical review of recent scholarship. *Sociology Compass*, 7(1):1–12.

Peer, L. and Ksiazek, T. B. (2011). Youtube and the challenge to journalism: new standards for news videos online. *Journalism Studies*, 12(1):45–63.

Petty, R. E. and Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology*, 19(C):123–205.

Philo, G. (1983). Bias in the media. *Socialist arguments*, 1:130.

Pons, P. and Latapy, M. (2006). Computing communities in large networks using random walks. *J. Graph Algorithms Appl.*, 10(2):191–218.

Pottker, H. (2003). News and its communicative quality: the inverted pyramid—when and why did it appear? *Journalism Studies*, 4(4):501–511.

Radicchi, F., Castellano, C., Cecconi, F., Loreto, V., and Parisi, D. (2004). Defining and identifying communities in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(9):2658–2663.

Rahimi, B. (2015). Censorship and the islamic republic: Two modes of regulatory measures for media in iran. *The Middle East Journal*, 69(3):358–378.

Reich, Z. (2011). User comments. In *Participatory Journalism*, pages 96–117. Wiley-Blackwell.

Robins, D. and Holmes, J. (2008). Aesthetics and credibility in web site design. *Inf. Process. Manage.*, 44(1):386–399.

Robinson, S. and Deshano, C. (2011). Citizen journalists and their third places: What makes people exchange information online (or not)? *Journalism Studies*, 12(5):642–657.

Rock, P. (1973). News as eternal recurrence. In *The Manufacture of News: A Reader*, pages 226–43. Sage Publications, London.

Rosen, J. (2008). A most useful definition of citizen journalism. *PressThink, July*, 14.

Ruiz, C., Domingo, D., Micó, J. L., Díaz-Noci, J., Meso, K., and Masip, P. (2011). Public sphere 2.0? the democratic qualities of citizen debates in online newspapers. *The International Journal of Press/Politics*, 16(4):463–487.

Salihefendic, A. (2015). How reddit ranking algorithms work. `https://medium.com/hacking-and-gonzo/how-reddit-ranking-algorithms-work-ef111e33d0d9`. Accessed: 2016-11-22.

Saltzis, K. (2012). Doing journalism online: How uk news organizations have adapted in the age of the internet. *The Handbook of Global Online Journalism*, pages 462–477.

Sanders, G. S. and Mullen, B. (1983). Accuracy in perceptions of consensus: Differential tendencies of people with majority and minority positions. *European Journal of Social Psychology*, 13(1):57–70.

Schlieder, C. and Yanenko, O. (2010). Spatio-temporal proximity and social distance: a confirmation framework for social reporting. In *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks*, pages 60–67. ACM.

Schmierbach, M. and Oeldorf-Hirsch, A. (2012). A little bird told me, so i didn't believe it: Twitter, credibility, and issue perceptions. *Communication Quarterly*, 60(3):317–337.

Scholz-Crane, A. (1998). Evaluating the future: A preliminary study of the process of how undergraduate students evaluate web sources. *Reference Services Review*, 26(3/4):53–60.

Schuler, D. and Namioka, A., editors (1993). *Participatory Design: Principles and Practices*. L. Erlbaum Associates Inc., Hillsdale, NJ, USA.

Schultz, J. (1998). *Reviving the fourth estate: Democracy, accountability and the media*. Cambridge University Press.

Schweiger, W. (2000). Media credibility—experience or image? a survey on the credibility of the world wide web in germany in comparison to other media. *European Journal of Communication*, 15(1):37–59.

Scott, J., Millard, D., and Leonard, P. (2015a). Citizen participation in news: An analysis of the landscape of online journalism. *Digital Journalism*, 3(5):737–758.

Scott, J., Millard, D., and Leonard, P. (2015b). Identifying similar opinions in news comments using a community detection algorithm. In *Social Informatics 2015*, pages 98–111. Springer.

Scott, J., Millard, D., and Leonard, P. (2016). Are open news systems credible? - an investigation into perceptions of participatory and citizen news. In *Proceedings of the 12th International Conference on Web Information Systems and Technologies*, pages 263–270.

Self, C. (1996). Credibility. *An integrated approach to communication theory and research*, 1:421–441.

Shao, G. (2009). Understanding the appeal of user-generated media: a uses and gratification perspective. *Internet Research*, 19(1):7–25.

Shapiro, I. (2010). Evaluating journalism. *Journalism Practice*, 4(2):143–162.

Sillence, E., Briggs, P., Harris, P. R., and Fishwick, L. (2007). How do patients evaluate and make use of online health information? *Social Science & Medicine*, 64(9):1853 – 1862.

Simons, M. (2007). *The content makers : understanding the media in Australia*. Penguin Books, Camberwell, Vic.

Singer, J. B. (2003). Who are these guys?: The online challenge to the notion of journalistic professionalism. *Journalism*, 4(2):139–163.

Slater, M. D. and Rouner, D. (1996). How message evaluation and source attributes may influence credibility assessment and belief change. *Journalism & Mass Communication Quarterly*, 73(4):974–991.

Spinuzzi, C. (2005). The methodology of participatory design. *Technical Communication*, 52(2):163–174.

Springer, N., Engelmann, I., and Pfaffinger, C. (2015). User comments: motives and inhibitors to write and read. *Information, Communication & Society*, 18(7):798–815.

Starbird, K. and Stamberger, J. (2010). Tweak the tweet : Leveraging microblogging proliferation with a prescriptive syntax to support citizen reporting. *Proceedings of the 7th international ISCRAM conference*, 1(May):1–5.

Stephens, M. (2007). *A history of news*. Oxford University Press, USA.

Strömbäck, J. (2005). In search of a standard: four models of democracy and their normative implications for journalism. *Journalism Studies*, 6(3):331–345.

Strömbäck, J. (2008). Four phases of mediatization: An analysis of the mediatization of politics. *The International Journal of Press/Politics*, 13(3):228–246.

Stromer-Galley, J. (2003). Diversity of political opinion on the internet: Users' perspectives. *Journal of Computer-Mediated Communication*, 8(3).

Sundar, S. S., Knobloch-Westerwick, S., and Hastall, M. R. (2007). News cues: Information scent and cognitive heuristics. *Journal of the American Society for Information Science and Technology*, 58(3):366–378.

Sundar, S. S. and Nass, C. (2001). Conceptualizing sources in online news. *Journal of Communication*, 51(1):52–72.

Sunstein, C. R. (2002a). The law of group polarization. *Journal of Political Philosophy*, 10(2):175–195.

Sunstein, C. R. (2002b). *Republic.Com*. Princeton University Press, Princeton, NJ, USA.

Tang, S., Blenn, N., Doerr, C., and Van Mieghem, P. (2011). Digging in the digg social news website. *IEEE Transactions on Multimedia*, 13(5):1163–1175.

Thorson, E. and Duffy, M. (2006). A needs-based theory of the revolution in news use and its implications for the newspaper business. Technical report, Reynolds Journalism Institute at the Missouri School of Journalism, Columbia, MO.

Thorson, K., Vraga, E., and Ekdale, B. (2010). Credibility in context: How uncivil online commentary affects news credibility. *Mass Communication and Society*, 13(3):289–313.

Thurman, N. and Hermida, A. (2010). Gotcha: How newsroom norms are shaping participatory journalism online. *Web journalism: A new form of citizenship*, pages 46–62.

Thurman, N. and Schifferes, S. (2012). The future of personalization at news websites: lessons from a longitudinal study. *Journalism Studies*, 13(5-6):775–790.

Thurman, N. and Walters, A. (2013). Live blogging–digital journalism's pivotal platform? a case study of the production, consumption, and form of live blogs at guardian.co.uk. *Digital Journalism*, 1(1):82–101.

Thurman, N. J. (2008). Forums for citizen journalists? adoption of user generated content initiatives by online news media. *New Media & Society*, 10(1):139–157.

Tsagkias, M., Weerkamp, W., and de Rijke, M. (2009). Predicting the volume of comments on online news stories. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, CIKM '09, pages 1765–1768, New York, NY, USA. ACM.

Tseng, S. and Fogg, B. (1999). Credibility and computing technology. *Communications of the ACM*, 42(5):39–44.

Tsur, O., Davidov, D., and Rappoport, A. (2010). A great catchy name: Semi-Supervised recognition of sarcastic sentences in online product reviews. In Hearst, M., Cohen, W., and Gosling, S., editors, *Proceedings of the Fourth International Conference on Weblogs and Social Media*. The AAAI Press, Menlo Park, California.

Väätäjä, H., Vainio, T., and Sirkkunen, E. (2012). Location-based crowdsourcing of hyperlocal news: dimensions of participation preferences. In *Proceedings of the 17th ACM international conference on Supporting group work*, pages 85–94. ACM.

Vis, F. (2013). Twitter as a reporting tool for breaking news. *Digital Journalism*, 1(1):27–47.

W3Techs (2015). Usage of content languages for websites. `https://w3techs.com/technologies/overview/content_language/all`. Accessed: 2016-08-11.

Wahl-Jorgensen, K., Sambrook, R., Berry, M., Moore, K., Bennett, L., Cable, J., Garcia-Blanco, I., Kidd, J., Dencik, L., and Hintz, A. (2013). Bbc breadth of opinion review: Content analysis.

Wanta, W., Golan, G., and Lee, C. (2004). Agenda setting and international news: Media influence on public perceptions of foreign nations. *Journalism & Mass Communication Quarterly*, 81(2):364–377.

Wathen, C. N. and Burkell, J. (2002). Believe it or not: Factors influencing credibility on the web. *Journal of the American society for information science and technology*, 53(2):134–144.

West, M. D. (1994). Validating a scale for the measurement of credibility: A covariance structure modeling approach. *Journalism Quarterly*, 71(1):159–68.

Whiteley, P., Clarke, H. D., Sanders, D., and Stewart, M. (2013). The economic and electoral consequences of austerity policies in britain. In *EPSA 2013 Annual General Conference Paper*, volume 138.

Williams, K. (2009). *Read all about it!: A history of the British newspaper*. Routledge.

Wilson, C. E. and Howard, D. M. (1978). Public perception of media accuracy. *Journalism Quarterly*.

Wojdynski, B. W. (2016). The deceptiveness of sponsored news articles: How readers recognize and perceive native advertising. *American Behavioral Scientist*.

Yadamsuren, B. and Erdelez, S. (2011). Online news reading behavior: From habitual reading to stumbling upon news. *Proceedings of the American Society for Information Science and Technology*, 48(1):1–10.

Yesil, B. (2014). Press censorship in turkey: Networks of state power, commercial pressures, and self-censorship. *Communication, Culture & Critique*, 7(2):154–173.

Zhou, D. X., Resnick, P., and Mei, Q. (2011). Classifying the political leaning of news articles and users from user votes. In *ICWSM*.