MEDIA@LSE
Department of Media and Communications

MEDIA@LSE MSc Dissertation Series

Compiled by Dr. Bart Cammaerts, Dr. Nick Anstead and Ruth Garland

Science in the headlines: the stakes in the Social Media Age

Sasjkia Otto,
MSc in Media, Communication and Development

Other dissertations of the series are available online here:

http://www.lse.ac.uk/media@lse/research/mediaWorkingPapers/ElectronicMScDissertationSeries.aspx

Dissertation submitted to the Department of Media and Communications, London School of Economics and Political Science, August 2013, in partial fulfilment of the requirements for the MSc in Media, Communication and Development. Supervised by Professor Robin Mansell

The Author can be contacted at: sasjkia.otto@gmail.com

Published by Media@LSE, London School of Economics and Political Science ('LSE'), Houghton Street, London WC2A 2AE. The LSE is a School of the University of London. It is a Charity and is incorporated in England as a company limited by guarantee under the Companies Act (Reg number 70527).

Science in the headlines: the stakes in the Social Media Age

Sasjkia Otto

> When fruit and vegetables are BAD for you: Getting your five-a-day is responsible for HALF of all food poisoning cases (Innes, 2013)

ABSTRACT

Neither the Leveson Inquiry nor the UK Parliament Inquiry into Science Reporting applied a social media lens to their investigations of science in the media. This paper employs content analysis of tweets linking to mainstream media articles on scientific research, and the corresponding articles, to consider the implications of current sharing patterns for the public understanding of science. It finds that actively participating audiences are not necessarily actively engaged, and headline content is crucial to the spread of information in the social media age.

INTRODUCTION

Science in the media has recently come under scrutiny—first in the Leveson Inquiry, and currently in the UK Parliament Inquiry into Science Reporting. The issues arising in both investigations are not new—with accuracy (including false balance) amidst political and economic pressures cited as main concerns about scientific information making its way from mainstream media to the public (Robin, 2013). Whilst Leveson (2012, p688) received evidence that 'science reporting had improved in recent years and that the majority of science reporting was responsible and accurate', the enquiry also heard that 'there is a tendency in parts of the press to sensationalise science news headlines' (p. 690). This is situated within the context of the Press Complaints Commission's approach to ruling on headlines only in the context of a story in its entirety (Beales, 2012), and its hesitance to arbitrate on scientific issues (Climate: Public Understanding, 2013).

However, there has been little attention by either investigation to contextual factors within the social media age, despite the fact that big media no longer has a monopoly over the news (Gillmor, 2009). There is currently much hype about audiences' participation in the publication process, and their influence over the spread of information. This links with scholarship in media and cultural studies on how audiences are active in making their own meanings from the texts they encounter (Hall, 2003; Livingstone, 2003), running parallel to sociology of science theories on the public's contextualisation and resistance of scientific information as it rides the juggernaut of second modernity (Beck, 1992; Locke, 2001), and the social shaping of technological outcomes (Wajcman and MacKenzie, 1999).

However, participation should not be confounded with engagement. There is currently little research on the extent and nature of engagement in meaning-making of participatory audiences. On the other hand, much policy focus on content assumes that audiences will receive information as sent. This paper moves towards bridging this gap in the context of online mainstream media coverage of scientific research. It employs content analysis of tweets linking to mainstream UK national publication articles, as well as the articles to which they link, to explore the research question:

What are the implications of ways in which Twitter users participate in the dissemination of UK mainstream online media coverage of scientific research?

This is examined through three sub questions:

1. In what ways are Twitter users engaging with traditional media coverage of scientific research?
2. What factors could be contributing to engagement patterns?
3. What are the implications of these engagement patterns?

The first sub question draws on media and cultural studies, and the sociology of science, to descriptively consider the ways in which audiences are creating meanings in the public domain, and critically engaging with information on science. Findings suggest that participatory audiences are not necessarily engaged with the content or with their peers. Whilst there is some suggestion of meanings being negotiated through personal relevance, most Twitter users promulgate headline content—with few references to other article text, and the small portion of personal comments indicating uncritical engagement. Despite these factors, there was limited utilisation of the space and features afforded by Twitter.

The second sub question uses generalised estimating equations (regression analysis accounting for correlations between clusters of observations; Rabe-Hesketh and Skrondal, 2008) to trace multiple sources of influence—from scientific authority, to mainstream media content, technologies, and different aspects of the network—to consider potential influences shaping audience engagement patterns. The data suggest that audience engagement cues could come from a combination of these (Foucault, 1978), but that headline content and social media sharing features of publications are particularly relevant. There is some indication that headlines containing advice have higher-than-average tweet volumes, and those containing generalisations with the potential to feed into stereotypes are more likely to be passed on verbatim. On the other hand, greater personal and critical engagement was found for headlines highlighting study limitations and conflicts with other information. Whilst horizontal interactions tended to have higher levels of engagement, these were in the minority, and the data suggest that the speed and ease offered by publication social media sharing features could contribute to engagement patterns.

The third sub question qualitatively examines the results of the first two—drawing on issues of veracity, ideology, and influence, and situating them within the current regulatory framework. The paper concludes that audiences cannot be expected to engage with all of the every-increasing flows of information they encounter online. Whilst some factors cannot be

controlled, adding a social media lens to consideration of mainstream media influence in scientific reporting would, firstly, increase emphasis on accuracy within headlines (and speed of corrections, where necessary) and, concurrently, employ various devices early in the article to promote engagement, as appropriate.

LITERATURE REVIEW

The widespread dissemination of blogging tools in the 1990s marked the beginnings of an 'ecosystem in which journalists, sources, readers, and viewers exchange information' (The People Formerly Known, 2011). With the integration of production and consumption, mass communication evolved into mass self-communication—as audiences become both receivers and senders of content (Castells, 2009), and both decoders and encoders of messages (Hall, 2003).

News as a shared social experience forms part of this participatory culture, with the potential for content transferred vertically from traditional publications to be disseminated and modified horizontally amongst citizens (Rosen, 2007). Eight out of ten American online news consumers received or shared links online by the beginning of 2010 (Purcell *et al.,* 2011), and about 13% of Twitter posts now contain URLs (Java *et al.,* 2013). This new type of producer-consumer has a particular interest in science and health-related information and the group looks set to grow, with online participators—especially the 19-29 age group—asking for more of this information above all other topics (Purcell *et al.,* 2011).

These factors depict audiences that are both interested and empowered to contribute to scientific information in the public domain, and all of the implied public goods (Stein, 2004). Studies of this new participatory culture synergise with social constructivist and social constructionist literature found across audience studies and the sociology of science (Locke, 2001), which—along with their poststructuralist cousins—distance themselves from notions of externally-determined social order (Kearney, 1994). Instead, they trace 'the origins of knowledge and meaning and the nature of reality to processes generated within human relationships' (Gergen and Gergen, p. 1).

Contrasting with previous strong media effects or reinforcement models (Scheufele, 1999, p. 105)—recent audience scholarship has focused on how people are 'active in shaping their media culture, contributing to the process of shaping and co-constructing their material and symbolic environment' (Livingstone, 2003, p. 4). Hall's (2003) model of encoding and

decoding outlines how audiences can be selective in the meanings derived from text—opting for the meaning dominant in the text, negotiating meaning based on personal relevance, or taking an oppositional stance. This last link in the communication chain can also be the first in a participatory cycle of production, circulation, reception, and reproduction—with meaning created at each stage, and contributing to the overall information in the public domain (Livingstone, 2003).

However, there are many different types of participation—not all of which are equally empowering to the participator (Arnstein, 1969)—and, some have argued, not all of which should be (Keen, 2008). Firstly, visible participation activities do not represent the majority—with as little as 1% of people who view content online considered content creators (Nielson, 2006)—although the 57% of American teens considered content creators could start to bridge the gap (Lenhart and Madden, 2005). Secondly, the act of participation should be separated from the process of meaning-making both personally and publicly, as it is possible for them to occur to different extents in the same context.

Moreover, the audience can only work with what is given (Bird, 2003). For information about scientific research, this can come from personal experiences, researchers and institutions, the media, and technology (Sundar and Nass, 2001)—all in the context of networks that are situated within specific structural frameworks. These factors all have potential power over how information spreads in the public domain. This paper's definition of power refines Castells' (2009) interpretation as 'the relational capacity that enables a social actor to influence asymmetrically the definitions of other social actor(s) in ways that favour the empowered actor's will, interest and values'; it adopts the notion of relational capacity whilst adding that influence is not always intentional or in favour of any one actor. To avoid claims about intention or benefit, this paper adopts the term influence, and considers it in line with Foucault's (1978) thesis on the diffuse nature of power.

Scientific influence

A 2010 survey found that most British people highly regard science (Public Attitudes to Science, 2011, pp2-7). Of the respondents, 88% and 82%, respectively, thought that 'scientists make a valuable contribution to society' and they 'want to make life better for the average person' (both increased since the previous survey in 2000). Moreover, 64% of respondents thought that 'experts and not the public should advise the Government about the implications of scientific developments'.

These figures lend some credence to the sociological deficit model on the public's understanding of science, which assumes that people are already persuaded about the value of science, which simply needs to accommodate their cognitive capacities (Gross, 1994). This links with perceptions that the public sees science as a rationalised image—'a universalistic, a-social monolith', with the ordinary person's mental state 'presumed to be a product of a rationalisation process of which science is both part product and producer' (Locke, 2001, p. 2). This monolith is perceived as casting a shadow over the PCC, which appears hesitant to arbitrate over scientific matters (Climate: Public Understanding, 2013).

The vision of the public as 'dupes of science' (Locke, 2001, p12) contrasts with Beck's (1992) treatment of reflexive scientization, which sees this transforming in the grips of second modernity. As the sciences become more complex, Beck argues, they become more subject to conflicting interpretations, and self-interested manipulation (Epstein, 1996)—especially in the context of risk (Beck, 1992). The public is accused of 'flip-flop' thinking for either mythologising or demonising science, as the need for science increases yet its claim to universal truth diminishes (Epstein, 1996). As such, Beck (1992, p. 155) sees the Enlightenment 'constellation of unbroken faith in science and progress' characteristic of the 20th century making way for either ambivalence or confrontation with the sciences; now, they are '…targeted not only as a source of solutions to problems, but also as a cause of problems' (p. 156).

Challenges to the validity of the scientific claims are neither historically nor presently unfounded—the imperialist appropriation of science for essentialising the colonial subject (Hall, 1997 and Said, 1978) and the 2009 Climatic Research Unit email controversy (Climategate, 2009) being cases in point. Not ignoring the dangers of asserting social factors over scientific knowledge (and, conversely, creating overly-high expectations of science though promoting a rationalised image), spheres of scientific enquiry tend to be subject to the economic and ideological requirements of power (Foucault, 1978). Rather than two distinct entities, science and society are deftly intertwined (Locke, 2001).

As audience theorists stress that personal experiences are relevant to meaning, so proponents of the contextual model in sociology see the public's understanding of science as a joint creation of scientific and local knowledge in a deliberative genre where social concerns are always relevant (Gross, 1994), and local knowledge is an important basis for opposition (Locke, 2001), as people actively manipulate scientific information to make it relevant to them. Dearing (1995, p. 343) notes that, 'unlike scientists, members of the general public often combine beliefs in science and superstition, with little apparent contradiction'.

These theories contrast with claims that the 1990s marked the 'end of ideology' (Thompson, 1990, p. 81) in the West, with the rise of pragmatism in decision-making. However, Thompson argues for 'mediazation' of the scholarship of paradigmatic shifts (Kuhn, 1968) to incorporate 'the ways in which symbolic forms in modern societies have become increasingly mediated by the mechanisms and institutions of mass communication' (p. 95).

Mainstream media influence

Most studies on media coverage of science have also found an overly-optimistic attitude—with failure to highlight limitations (O'Connor and Rees, 2012), risks, and ethical shortcomings (Condit, 2004). Journalists have been accused of providing a public-relations service for science (Cheerleader or Watchdog, 2009), which could be partly attributed to the need to cultivate sources (Hansen, 1994).

This relationship between science and the media has a long history. The construction of the Orient would scarcely have been possible without the diffusion of racialised representations in popular culture (Said, 1979, and Hall, 1997). This relationship refracts as scientific knowledge moves through the prism of biases associated with publication for public consumption, and is rarely transplanted intact into the public domain (O'Connor and Rees, 2012). Carvalho (2007, p232) found that information on climate science undergoes 'strong ideological filters' in different publications, with accuracy secondary to political leaning. O'Connor and Rees (2012) further found that neuroscience in the press was used to essentialise identity by gender and other traits, and justify existing family and labour practices—sometimes extrapolating beyond the research.

> As science penetrates the public sphere, it enters a dense network of cultural meanings and worldviews and is understood through the prism they provide. The cultural context determines which aspects of science travel into public consciousness: knowledge that resonates with prevailing social norms is selectively 'taken up' in public dialogue. (O'Connor and Rees, 2012, p. 220).

These factors, in turn, feed back to the media's support for the 'social authority and power of science as the guardian of truth' (Carvalho, 2007, p. 225), which has implications for the public's construction of both science and life-worlds.

It is not uncommon for inaccuracies to make their way into publications. For instance, a recent headline claiming 'Racism is Hardwired into the Brain' has since been refuted by the

New York University researchers concerned (Leveson, 2012, p. 691). Cooper et al. (2011, p. 1) found that between 68% and 72% of dietary claims in the UK's largest 10 newspapers 'had levels of evidence lower than the convincing or probable categories that are recommended for dietary health claims'. Perhaps the most infamous example of the UK media's distortion of scientific consensus is the scare linking the measles, mumps, and rubella (MMR) vaccine to autism, which—four years following the publication of the small research piece in question—saw less than a third of broadsheet reports on the topic referring to the 'overwhelming evidence' that the MMR vaccine is safe (Goldacre, 2008).

Such media shortcomings challenge Keen's (2008, p. 15) remark about a dethroned 'dictatorship of expertise', as the media has proven capable of misinformation well in advance of the social media boom. Notwithstanding this, Leveson (2012, p. 693) expresses concern about content quality amidst 'the commercial and competitive pressures operating on journalists in an extremely competitive market'—despite coverage of science being one of the most consistent sources of profit (Robbins, 2011). For instance, 46% of respondents to Fox's (2010) survey of UK science correspondents indicated that time available for fact-checking had reduced. Moreover, Leveson (2012, p. 480) reports that '…a considerable number of witnesses and commentators have complained about the use of misleading and inaccurate headlines [across all topics], often it seems knowingly used in order to attract custom' (p480).

There is, however, the risk of confounding legitimacy with influence—as the state of the former does not necessarily say anything about the latter, although the literature does provide some clues. Thorson (2008) found that articles containing advice were more likely than others to remain on the New York Times website's most emailed list for several days. Moreover, a Cochrane Collaboration meta-analysis of studies on the effects of mass media in the utilisation of health services suggests that presentation of scientific information in unplanned media coverage is not without consequence (Grilli *et al.,* 2009), and Leveson (2012) presents the drop in MMR vaccination rates and accompanying spike in measles cases as a case in point.

There has been extensive research on the ways in which the media can shape agendas and norms (Curran, 2002). Couldry (2001) discusses the media's symbolic authority over the ordinary for constructing understanding of life-worlds. This proposition is underlined by public media consumption in the UK: most people cite some form of traditional media as a regular source of science information both online (Horrigan, 2006) and offline, whilst only 2% say that they turn to blogs (Public Attitudes to Science, 2011).

Despite this primacy of traditional news sources, however, audiences report that they are less trusting of these than scientific journals, but feel obliged to turn to the former due to perceived inaccessibility of the latter. Moreover, whilst a small majority (47%) agree that the information they hear about science 'is generally true', 34% say they are undecided, and 9% disagree. That said, people also say that they find information more trustworthy when it has been checked by third parties, including journalists (Public Attitudes to Science, 2011, p. 3).

Technological influence

Research suggest that social networking tools rank lower in credibility perceptions than mainstream media—although the latter can serve to augment the former's clout (Morris et al., 2012). One experiment suggests that people find headlines presented on traditional news sites more credible than the same ones on Twitter (Castillo *et al.,* 2011). However, regular Internet users may find online sources in general more trustworthy (Castillo *et al.,* 2011), and views may evolve as the Twitter population grows (Brenner and Smith, 2013).

A more significant influence could be attributed to the technologies' technical specifications (Lessig, 2006). Microblogging sites like Twitter are praised for freeing people of the 'abstraction and commitment' of composed blogs (Nutall, 2007), and filling the need for faster communication (Java et al., 2007). More cautiously, Robbins (2012a) asks: 'Social media didn't exist at the time of the main MMR scare, but if it had, what influence would it have wielded?'

These features, moreover, have the potential to constrain the access and processing of information. Winner (1986) argues that technologies are not neutral, with their designs setting parameters. For Twitter, a solid barrier to creation and engagement (Jenkins *et al.,* 2009) is raised by the 140-character limit. Twitter's mobile-friendly features and non-obvious display of responses to posts all have the potential to shape user posts in more subtle ways. Social media sharing facilities embedded in mainstream news websites also facilitate sharing: Clicking the embedded 'Tweet' will render the full headline, and the name of the publication (and sometimes the author and section names) ready for dissemination on Guardian.co.uk, Telegraph.co.uk, and Independent.co.uk, whilst Dailymail.co.uk renders an abridged version (sometimes incorporating other content).

A combination of such fast-diffusion features, which add to the expanding proliferation of information found online, and to the burden of filtering and processing, have been blamed for reduced attention to content (Manjoo, 2013). Fox (2006) found that three-

quarters of people who search online for health information do not check the source and date, whilst web analytics company Chartbeat found that 38% of website visitors do not scroll past an article's headline, and share the link before scrolling beyond the third paragraph (Manjoo, 2012). It is unclear whether these patterns are related to sharing button positions, which are toward the top of the articles on Guardian.co.uk, Independent.co.uk, and Telegraph.co.uk, and below the article on Dailymail.co.uk.

Livingstone (2003) asks whether certain technologies can shape certain audiences, or social relations shape communicative possibilities. Diffusion of innovation theories (Rogers, 1995) have been criticised of oversimplified determinism—spawning an opposing body of literature within the social shaping of technology tradition (Wajcman and MacKenzie, 1999). An important contribution to these agency-centred approaches is Pinch and Bijker's (1984) social construction of technology theory, which argues that technological artefacts are underdetermined products of intergroup negotiations. 'Design ceases not because the artefact works in some objective sense but because the set of relevant social groups accepts that it works for them.' (Klein and Kleinman, 2002, p. 30).

However, such theories have also been criticised for neglecting non-social influences.

> On the one hand, diffusion scholarship has tended to overlook the degree to which media artefacts is tied to their social construction. On the other hand, social shaping research has largely neglected the extent to which the development of artefacts is linked to their planned and actual diffusion. (Boczkowski, 2004, p. 255).

Boczkowski's (2004) analysis of videotext newspapers, for example, found that partial technological outcomes had the power to shape the social and vice-versa—in line with recent scholarship attempting to move beyond this duality, which instead emphases the interdependence of technological and social transformations, the ongoing character of the process, and the influence of the historical context in which it unfolds. Jenkins' (2009, p. 6) summarises this in his assessment of participation:

> Some tasks may be easier with some technologies than with others, and thus the introduction of a new technology may inspire certain uses. Yet these activities become widespread only if the culture supports them, if they fill recurring needs at a particular historical juncture. The tools available to a culture matter, but what a culture chooses to do with those tools matters more.

Network influence

Issues of mistrust in social media messages may be attributed partly to the uncontrolled nature of microblogging, which makes it vulnerable to credulous users (Ravikumar *et al.,* 2012) and, as such could be somewhat mitigated by the disproportionate weight that people assign to information coming from peers as sources (Sundar and Nass, 2001).

The network's influence stems from its ability to carry messages between people (Moody, 2010), and information has the potential to spread in unprecedented ways as people across demographics connect horizontally and switch amongst multiple networks (Wellman, 2001). Many discussions on social media are using phrases such as 'wisdom of the crowds' as institutions, people, and machines come together to filter information (Gruber, 2008). On the other hand, Hindman (2009) found that most online talk is in the hands of the few, with highly-visible sources becoming even more so. Moreover, Rafaeli and Sudweeks (1997) found that only about a tenth of computer-mediated communication groups contain interactive comments between users.

This spread and filtering of information could be associated with the tone of the message. Muchnik et al. (2013), for instance, found a herding effect around positive messages, whilst Berger (2013) found that good news spreads faster on social media. These factors have to potential to interact with other societal processes, and mitigate public displays of mistrust and confrontation that may be present offline (Beck, 1992).

(Self-)Regulatory influence

All of these influences can be augmented or suppressed by the overarching regulatory framework, which in the UK is predominantly embodied by the voluntary PCC. Particularly relevant to social media is the PCC editorial code stance on headlines, which are not covered in isolation:

> The accuracy clause particularly—with its requirement to take care not to mislead or distort—applies to them, but taken in the round and set in the context of the story as a whole. So the Code, and the PCC's interpretation of it, gives headline writers ample latitude to produce eye-catchingly baited hooks to tempt the reader (Beales, 2012, p. 26).

One complaint to the PCC concerned a Telegraph headline containing an unsubstantiated claim by a teenager's parents that the vaccine for human papillomavirus (linked to cervical cancer) left her in a 'waking coma'. The PCC ruled:

> …due to their necessary brevity, headlines can represent only a limited summary of a potentially complex set of circumstances… While the Commission acknowledged that the headline itself did not clearly denote the assertion as representing a claim, the sub-headline clearly stated that 'the parents of a 13-year-old girl believe the cervical cancer jab has left their daughter in what they describe as a 'waking coma'… The newspaper was fully entitled to report the concerns raised by the parents, provided that it clearly distinguished their conjecture from fact. The Commission was satisfied that the newspaper had complied with this requirement by establishing this in the sub-headline and as such did not consider that readers would be significantly misled. (PCC Judgment on Daily, 2012, comments section).

Another judgment on the same story in the Daily Mail ruled—two months after the initial complaint—that quotation marks should be added to the headline to indicate that the claim is an opinion. This contrasts with Leveson's (2012, p693) recommendation that any new regulator should 'bear… closely in mind' guidelines including that headlines '…should not mislead the reader about a story's contents and quotation marks should not be used to dress up overstatement' (Robbins, 2012a).

Audience influence

All of these factors link back to audiences by helping to shape the diversity of sources and content available (Napoli, 1999). Audiences, in turn, help to structure the spread of scientific information online—through participation even more so now than in the age of voting with their wallets (Thorson, 2008). Tweeting links can serve as public endorsement (or criticism) for any number of factors, and help users filter and aggregate information—with social networking sites beginning to rival some search engines as the source of visitors to big news sites (The People Formerly Known, 2011).

However, research is thin on the current outcomes of audience participation in processing and spreading scientific information online. Information processing patterns have implications for the personal and public understanding of science. A diversity of access to sources and content does not necessarily equate to a diversity of access to exposure—defined as citizen engagement with a variety of content available to '…increase their knowledge, encounter opposing viewpoints, and become well-informed decision-makers who are better capable of fulfilling their democratic responsibilities in a self-governing society.' Simply, it refers to content 'as received'—which now has unprecedented implications for content 'as sent' (Napoli, 1999, p. 24).

CONCEPTUAL FRAMEWORK AND RESEARCH OBJECTIVES

'Interactive media hold the potential to transform a once-mass audience into engaged and participatory users of information and communication technologies' (Livingstone, 2003, p. 27). Much literature has focused on a shift of influence in favour of the audience (Gillmor, 2009)—for better or worse (Keen, 2008). However, many aspects of such a shift are yet to be systematically investigated (van der Graaf, 2009) with due focus on context (Livingstone, 2003). To date, there is little data considering the extent to which participatory audiences are actually engaged with the content and each other.

This paper moves beyond the traditional focus on text content, to also consider relationships with audiences (Napoli, 1999). It considers various aspects of Twitter audience interaction with publications and other Twitter users by examining multiple dimensions related to tweets and the articles to which they link—particularly focusing on headlines. It considers how determinist and social constructivist/constructionist perspectives could mutually contribute to understanding the 'multiple overlapping and intersecting sociospatial networks of power' by tracing various sources of influence potentially shaping information spread on Twitter (Mann, 1986, p. 1).

Through the act of sharing a link to an article online, the audience is already active in the participatory sense. However, different patterns may be observed in both the volume and nature of engagement as related to content processing and interaction activities. With new participatory technologies, '…reception may be once again gleaned—at least to some extent—from an analysis of use.' (Livingstone, 2003, p. 26). There are potential restraints on publishing decisions due to unprecedented visibility to various parties (Ellerbrok, 2010)—especially on a predominantly public service such as Twitter. For this, and practical reasons, the decoding process will unlikely be fully reflected in the content that users publish. However, this public decoding process could informative and of consequence to the personal and public understanding of science.

Within this framework, this paper situates its central research question:

What are the implications of ways in which Twitter users participate in the dissemination of UK mainstream online media coverage of scientific research?

Treatment draws Jenkins et al.'s (2009) notion that participation does not have guaranteed outcomes, and it is necessary to foster means of deploying potentially useful characteristics to useful ends.

Three sub questions are deployed:
In what ways are Twitter users engaging with traditional media coverage of scientific research?

This question examines trends in nature of engagement, with attention to theories on active audiences and reflexive scientization.

What factors could be contributing to engagement patterns?

This question attempts to glean what Tweets can reveal about the importance attached to and decoding of different types of science articles, and how communication patterns relate to article characteristics and technological functions—using different lenses in line with sources of influence identified in the literature.

What are the implications of these engagement patterns?

This question qualitatively assesses the findings in the first two. It draws on the issues identified in the role that science plays in society in terms of providing templates for behaviours and ways of thinking. It then considers potential ways of dealing with engagement practices in terms of article content (Grillli *et al.,* 2009). This does not seek a single best practice, but rather devices that can be employed to promote optimal engagement with relevant messages under different circumstances, and is situated in the context of the current regulatory framework, and issues investigated in the Parliament Select Committee on the Public Understanding of Science (2013).

RESEARCH DESIGN AND METHODOLOGY

This study employs content analysis of tweets containing links to articles on scientific research, and of the articles themselves (focusing on headlines)—exploring various variables with the potential to shape outcomes Scheufele's (1999). First, tweet characteristics are considered in isolation, to glean the prevalent practices in the decoding process. Subsequently, different sets of explanatory and response variables consider factors that could be influencing the types of information published, and the ways in which it spreads—treating both article and tweet characteristics as explanatory variables. Finally, article characteristics are considered on their own, to gauge potential implications of publishing patterns in terms of the relationships found.

This approach appropriates content published on social media sites as an alternative to the traditional use of surveys to study audience frames as a dependent variable, to consider a circle of inputs, processes, and outcomes—where outcomes of some processes serve as inputs for others (Scheufele, 1999). The method is considered best-placed to gauge general trends in participation in the dissemination of science articles, and look for potential patterns in the relationships between different variables (Hansen *et al.,* 1998).

Sampling

All relevant articles were retrieved over four weeks from June 22, 2013 to July 19, 2013, with the criteria that the article:

- Is based on a specific scientific study with a human response variable—either physical or nonphysical—conducted by any party/ies regardless of affiliations or channel(s) of presentation.
- Is published on Dailymail.co.uk, Guardian.co.uk, Telegraph.co.uk, or Independent.co.uk.
- Appears in the science/technology or health section(s).

Focus is on research because it represents up to 70% of weekly science stories, and sets much of the science agenda (Fox, 2001). The human focus looks for relevance to personal life-worlds. Mainstream media publications were chosen because of their primacy as online news sources (Horrigan, 2006), and the specific publications because they are the UK's largest for circulation and social media impact (Ponsford, 2013, and How are UK National, 2011) (see appendix A).

Due to the nonchronological nature of article presentation on Dailymail.co.uk, the LexisNexis archive was searched for articles within the topics 'Science & technology', 'Food science &

Technology', 'Health Care', and 'Prevention & Wellness'. Articles rendered were selected as per the criteria. This yielded 138 articles: 59 from Dailymail.co.uk, 46 from Telegraph.co.uk, 26 from Guardian.co.uk, and seven from Indpendent.co.uk. Due to inconsistent metatagging on these websites, and the LexisNexis archive, it is possible that some articles meeting the first two criteria were excluded from the sample due to the third.

The sample is limited to a four-week period because Twitter currently sets a seven-day limit for retrieving all relevant tweets (Hawksey, 2013). Due to the timing during Britain's traditional 'silly season' (Krippendorff, 2004) and summer holiday period, and other factors potentially contributing to the type of research published, these articles are not considered fully representative of the annual population of articles meeting the criteria.

Twitter Archiving Google Spreadsheet Tags v5—an open-source automatic tweet archiving platform—was used weekly to retrieve tweets linking to sample articles, with tweets retrieved up to four weeks following article publication (Hawksey, 2013). Search terms included full links and abbreviated versions rendered by publication tweet buttons and bitly.com. This yielded 11,775 tweets, for every fifth was systematically selected via macro—yielding a 20% sample stratified by publication (Bauer, 2000): 887 from Guardian.co.uk (the highest at 34.12 per article), 688 from Dailymail.co.uk (11.66 per article), 659 from Telegraph.co.uk (14.33 per article), and 121 from Independent.co.uk (17.29 per article) (see appendix B for a list of all articles, and the first tweet coded for each). Whilst a minority of sophisticated browsers may use other article linking services, the archiving tool was sensitive to a variety of linking conventions, and was cross-tested with professional web analytics software used on a trial basis (Topsy Pro Analytics, 2013). The sample is thus considered representative of the tweets linking to the sample articles by publication (Krippendorff, 2013).

The coding frame is based on immersion in the literature and sample texts. It considers 23 article and 16 tweet variables (with one qualitative each)—some automatically generated by the archiving tool (see appendix C).

Variables

Three different categories of variables were coded—for tweets, and article headlines/leads and bodies, respectively.

Tweet variables

The total number of tweets linked to articles was coded to gauge engagement volume. Other variables look for nature and extent of user engagement. They consider, firstly, whether the tweet contains the full/partial headline and/or other content and, secondly, whether it refers to article content outside the headline—with non-headline text taken is clues to higher levels of engagement, and headline text to the weight attached to such content.

Thirdly, tweets were coded for personal comments and, fourthly, for whether they could be considered uncritical—with tweets without personal comments automatically included in this category, and those containing criticisms of the article or the study, drawing comparisons to other information, assessing implications (beyond what is said by the article), or attempting to engage others' opinions (and any other suggestions against uncritical reception) excluded. The uncritical category is defined narrowly, for sensitivity to suggestions that information presented is not taken at face value (hence, the counterintuitive variable name). Whilst this variable does not distinguish between reception of the article and research, it hints at engagement level, and diverging interpretations of texts (Beck, 1992; Livingstone, 2003).

Tweets containing personal comments were also assessed qualitatively through an exhaustive list of categories within the uncritical and other values. Only primary categories were coded, as few tweets contained indication of secondary categories. The text revealed little user distinction between the article and the research, which is reflected by the variable categories. The categories are not mutually exclusive and were not treated statistically (Bauer, 2000), but add colour to the decoding process picture (Hall, 2003), depicting different strategies people employ when filling gaps or reframing meaning when commenting on information, and finding suggestions of information being appropriated in line with people's tacit knowledge (Livingstone, 2003).

Tweet date variables consider publication-to-publication timeframes, and engagement at different times of the week.

The flow of information and extent of vertical and horizontal interactivity is assessed through whether the tweet was marked as a retweet—from a publication account or otherwise—and whether it mentions other users. These variables investigate relationships with different sources (Sundar and Nass, 2001), and engagement for different types of interaction.

Tweet variables relating to technological influences include whether the tweet came from a mobile device, to consider whether mobility is associated with reduced engagement, and the tweet character count, to see how people are making use of the space afforded.

Article headline/lead variables

The main article variables focus on headline characteristics, due to their centrality to the analysis in the context of social media sharing features and the current (self-) regulatory framework. The same variables are applied to leads as secondary to consider relative influence.

Research suggests that text marked as controversial is processed differently (Epstein, 1996), framing statement as agreements could increase chances of acceptance (Kim and Rudin, 2013), and counterintuitive information can contribute to sharing patterns (Thorson, 2008). Coding therefore considers whether the text highlights study limitations, and contradictions with other information. Framing as policy recommendation is considered to see whether issues affecting society as a whole could generate different types of responses (Beck, 1992; Thorson, 2008).

Other headline variables consider content relating to issues with science identified in the literature (Said, 1978; Hall, 1997; Cooper *et al.,* 2011) and identified differences in sharing patterns (Thorson, 2008)—including whether the text generalises about groups of people in a manner that could give rise to stereotypes, or contains information that could be construed as advice. Whilst not all variables of framing per se (Scheufele and Tewksbury, 2006), they are designed to gauge sharing patterns for text containing these types of information.

Based on findings that tactical phrases can resonate with people (Luntz, 2013), coding identifies the word 'science' or any others derived from it as root.

A headline character count variable assesses interaction with technological constraints.

The type of research response variable (physical-only or otherwise), and whether a single person could affect the outcome of results reported explores potential variation in sharing patterns. Only the headline and lead were considered, as interpretations tend to depend on order of information presentation (Bergus *et al.,* 1998).

A qualitative variable identified all articles covered by the NHS Behind the Headlines (2013) fact-checking service, to consider potential implications of accuracy issues—classifying the headlines (and leads) as either inaccurate or misleading. As this service covered only 19 of the variables in the sample, it serves an anecdotally descriptive purpose only.

Article body variables

Article body variables are derived from literature on factors contributing to the perception of science, and the potential to promote different types of engagement—exploring decision-rules employed for engagement (Sundar and Nass, 2001).

Style-related variables include the Fog Index readability measure, and word count for clues on ease of engagement (Gunning, 1969).

Sourcing influence is investigated from multiple angles (Sundar and Nass, 2001). For perceptions of scientific authority, variables consider whether the article was written by a science specialist (Fox, 2010), quotes study sources, and what are research party affiliations. Other sources quoted were coded for indication of diversity (Napoli, 1999). All sources are coded for sex, to account for stereotypical views of scientists as male (Public Attitudes to Science, 2011). Publication impact could be due to sourcing or other factors, including readership/social media impact, user interface, and political leaning—which are all considered under the publication variable.

Intercoder reliability

A second coder was trained to test reliability for 20 articles and 70 tweets. The qualitative variables were tested through calculating intracoder reliability for 73 tweets and 19 articles, one week after initial coding, which is considered acceptable when a second coder is not available (Schreier, 2012). Reliability was calculated with ReCal online software (Freelon, 2013) using Krippendorph's Alpha—chosen for its ability to account for chance agreement, and suitability for different measurement levels (Lombard *et al.,* 2002).

Intercoder reliability was above 0.7 for all variables, which is considered acceptable for exploratory research (Lombard *et al.,* 2004). However, the sample of articles tested was smaller than the 30 generally recommended—although still larger than 10% of the sample frame (Riffe *et al.,* 2005)—and some tweets contained duplicate wording, with few containing personal comments. A larger sample needs to be tested for the coding frame to be considered robust. Intracoder reliability for the qualitative variables was above 0.8. Whilst this is less reliable than intercoder reliability, it is considered sufficient due to secondary importance to analysis (Schreier, 2012).

Analysis

As the articles sample is not considered fully representative, article-specific analysis is confined to descriptive statistics (Agresti and Finlay, 2009). Some tweet analysis also employs descriptive statistics.

Binary logistic regression analysis considers the relationships of various explanatory variables to the tweet response variables—adopting the $p<=0.05$ significance level, to generalise to the sample frame of tweets linking to the articles studied. Models were derived by first looking for significant associations between explanatory and response variables in isolation, then combining the significant explanatory variables for each response variable into a model, then removing one-by-one insignificant variables until none remained. Interactions between article controversy and source variables—exploring relationships between readers' opinions on sources and their opinions on messages when the information runs contrary to their beliefs (Sternthal *et al.,* 1978)—were not significant.

To account for potential violations of the assumption of observational independence (Agresit and Finlay, 2009), numbers reported are from generalised estimating equations which estimate 'marginal or population-averaged effects taking into account the dependence among units nested in clusters' (Rabe-Hesketh and Skrondal, 2008, p. 273). Exchangeable working correlations were specified for article-specific clusters in SPSS, to accommodate different cluster sizes.

Limitations

Whilst all stages of the study were carefully constructed, timing and resources limited some aspects of the research and subsequent analysis.

Coding and analysis

- Content analysis constructs a simplified version of reality that skims over context (Bauer, 2000). However, it remains useful as indicator of 'general tendencies and likely outcomes' (Napoli, 1999, p. 29), and care was taken in selecting various explanatory variables. Nonetheless, further studies may take a more subject-specific, qualitative approach.
- As the assumption is that audience effects generally occur with substantial attention to texts (Scheufele and Tewksbury, 2006), article-specific variables should be treated

cautiously—in light of the potentially weak relationship between scrolling and sharing (Manjoo, 2013), and the possibility that retweet authors did not visit the article page. Whilst analysis did account for clustering, future studies may also treat publication account tweets as explanatory variables.

- User attributes, which could add explanatory value (Scheufele and Tewksbury, 2006; Bird, 2003), was not included in the analysis. More comprehensive modelling may account for this in future.
- This analysis could not consider quantitatively article veracity. Cooper et al. (2011) agreed to share data from their study on the validity of dietary claims in the UK national press. However, the 2008 sample yielded no Twitter impact when searching historical data with professional web analytics software (Topsy Pro Analytics, 2013). This limits judgment on audience interpretive decisions.

Sampling

- Facebook and Twitter, in 2011, were used by about 50% and 9% of the UK population, respectively; however, Twitter's influence is rising—especially as most tweets are open to the public and search engines (The People Formerly Known, 2011). In light of this minimal expectation of privacy, Twitter was also deemed more ethical for analysis (Mareck, 2011).
- As Mailonline.co.uk is the only website that indicates if and when an article was updated since publication, coding could not account for post-tweet article amendments. Future studies may consider retrieving and coding on a daily, rather than weekly, basis.
- Replies to the tweets, which may not include the links, were not captured by sampling. Whilst these represented a minority on Twitter, they may reveal different engagement patterns, which could be explored in future studies.

ANALYSIS

In what ways are Twitter users engaging with traditional media coverage of scientific research?

This sub question examined what/how people are tweeting, when they are tweeting, and from what/to where they are tweeting.

What and how are people tweeting?

The average total number of tweets per article was 85.

Most of the tweets (73%) contained the headline in some form, and nearly half (43%) did not contain any content apart from the article headline (see chart 1). Similarly, 70% did not refer to article content outside the headline, and 17% referred to the lead only—leaving only 13% referring to article content outside the headline and lead (see chart 2). Public engagement

Chart 1: Tweet content characteristics

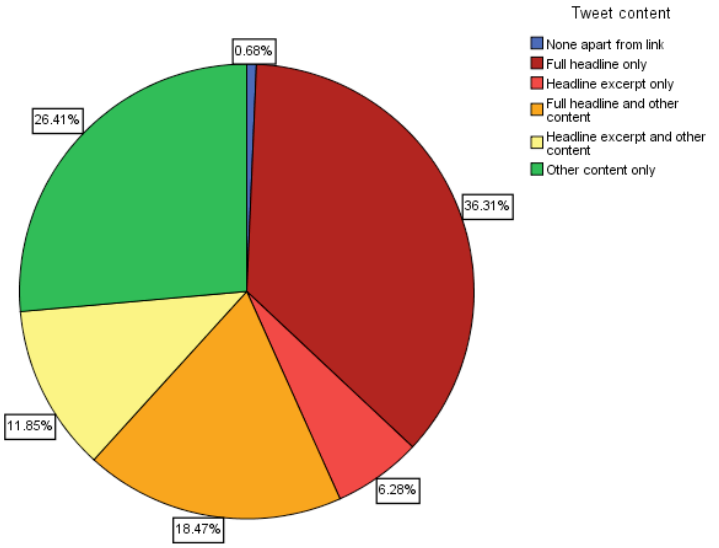
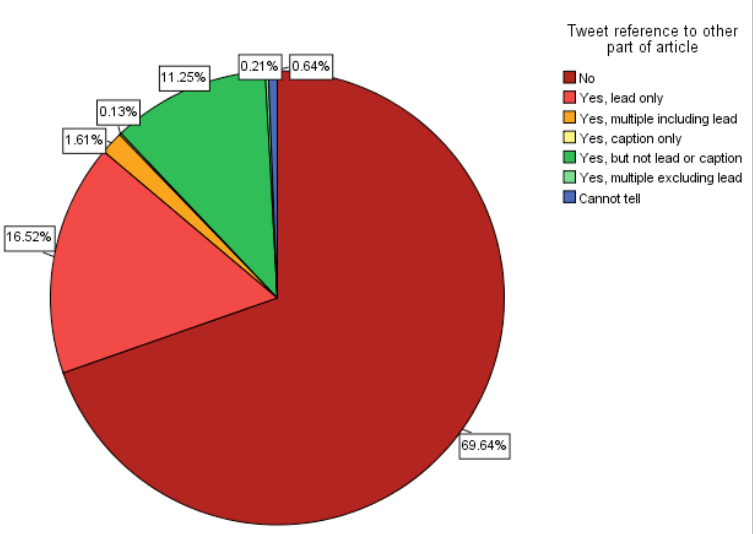


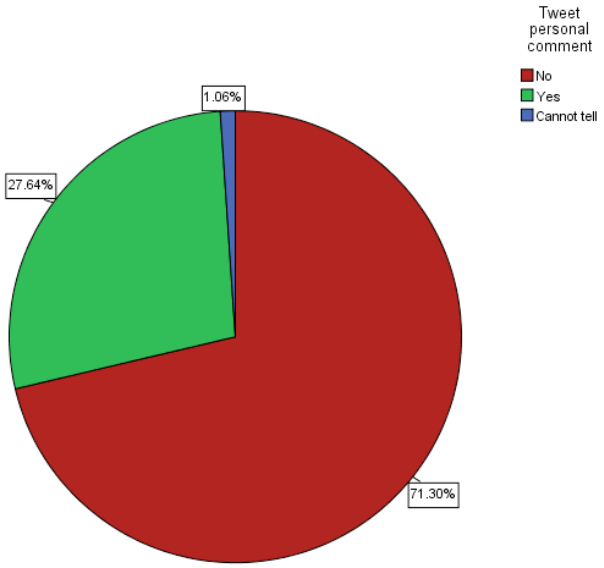
Chart 2: Tweets referring to article content outside the headline



with the articles thus appears focus on headlines.

Whilst 57% of tweets contained content other than the headline, only 28% contained a personal comment from the user (see chart 3).

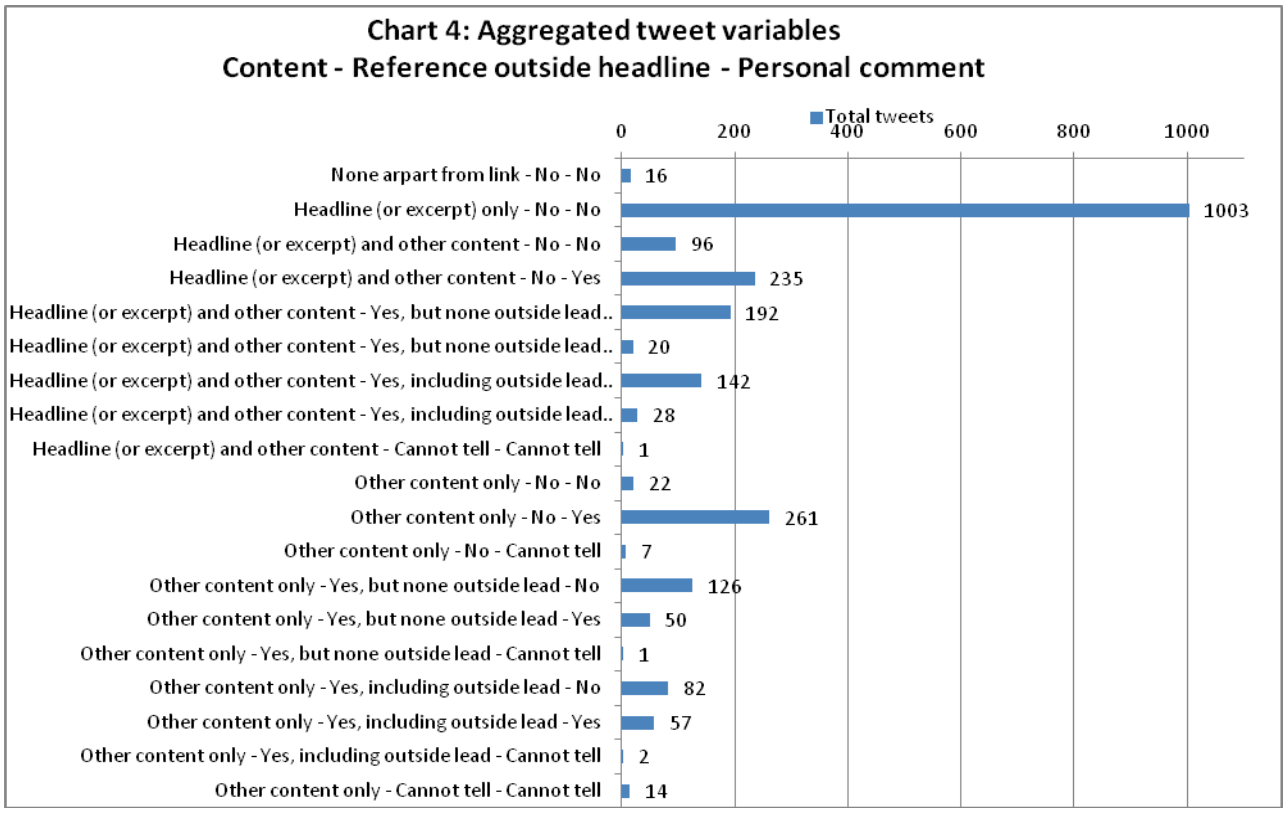
Chart 3: Tweets containing personal comments from the user



Notably, for all three the primary measures of engagement employed, roughly a quarter displayed higher levels of engagement coded (tweet content does not include the headline, refers to article content outside the headline and lead, and contains personal comment). Aggregating these variables further indicates that some content other than the headline may just be paraphrasing the headline—as indicated by tweets containing other content but no

comments or references to other parts of the article.

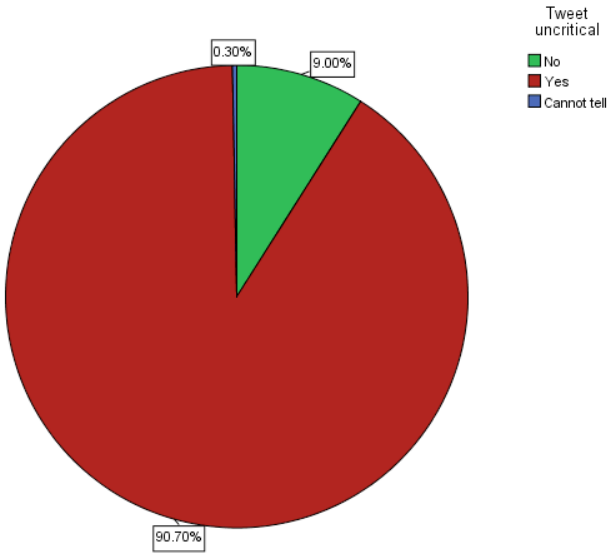
Chart 4 shows 118 tweets in this category and, in total, 48% contained of all tweets neither a personal comment nor reference to article content outside the headline. A further 14% contained no reference to content outside the headline and lead, and no personal comment. As such, only 38% of tweets contained any content—from the user, the article, or both—that



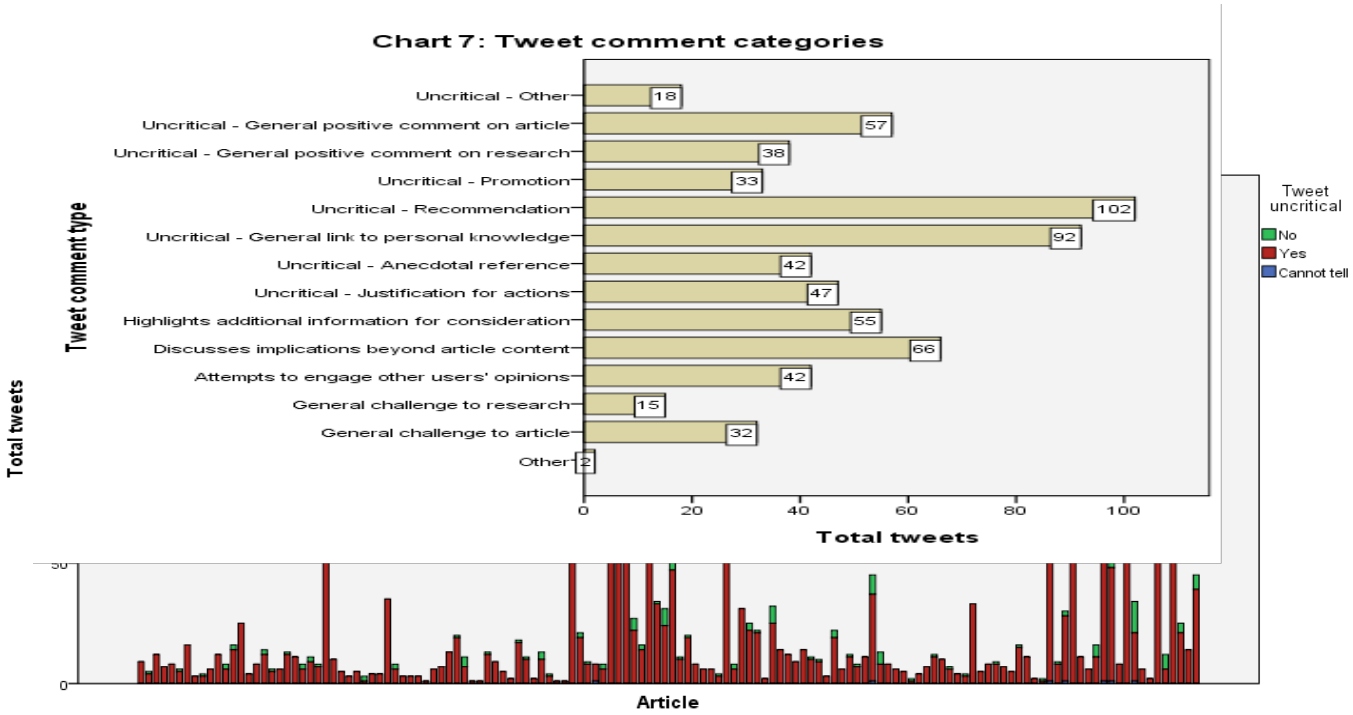
could not be found in the headline or lead (barring descriptive content such as publication title and section, which were excluded from coding).

An even smaller percentage (9%) did not indicate an uncritical reception of information presented (see chart 5). However, there is indication of a divergence of receptions, as this category is spread across articles (see chart 6).

Chart 5: Tweet indicates uncritical acceptance of article



When looking at the tweets that contain personal comments alone, those classified as uncritical (492) total more than double the number of those that are not (212) (see **chart 7**).



The more qualitative breakdown of their content does, however, indicate a heterogeneous set of interpretations, with a concentration around the higher-engagement categories depicted at the centre as opposed to lower-engagement categories toward the ends (such as outright acceptance or rejection of the articles or research without context). This indicates that, of the users that are engaging beyond the article's content, most are finding a way to make it relevant to what they know.

Overall, however, most tweets simply reflect the linked articles—particularly the headlines (and leads). These engagement patterns indicate that most fall within what Hall (2003, p. 56) would call the 'dominant' or 'preferred' code, with little evidence of exposure diversity beyond the headlines (and leads) (Napoli, 1999). However, where there is engagement, the data suggest that people draw on heterogeneous information that is relevant to them when assessing the content (Livingstone, 2003)—although only a minority provide public evidence of doing so. Linking with Beck's (1992) theories of primary and reflexive scientization, the data provide some evidence of conflicting views—even for the same texts (Livingstone, 2003)—but most reception was classified as uncritical, and the processes of symmetry implied by the contextual model (Gross, 1994) do not appear to apply in this context (regardless of what may be happening offline). Twitter users may be using complex processes

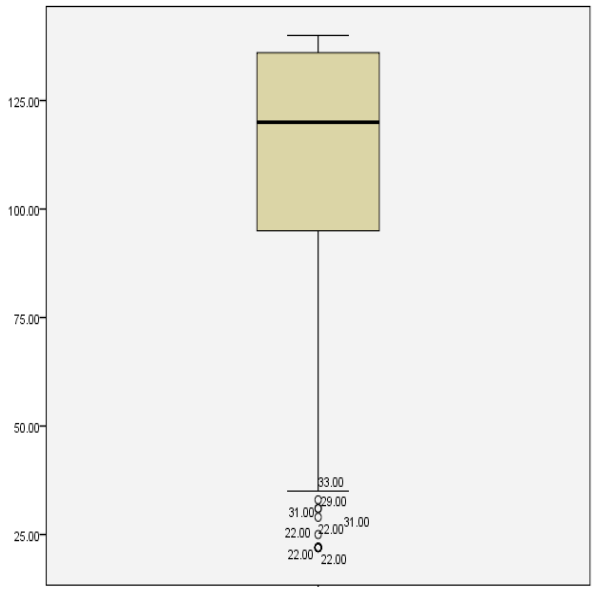
to make sense of their lives in a world where various factors are beyond their control (Mosco, 1996), but the weight of the data do not show public meanings made as equally within and against the resources provided by science and the media (Morley, 1993).

Condit (1999, p. 178) writes that

> ...the public is not a dupe to anyone—not the media nor scientists nor politicians. Recognizing that each of us can make intelligent interpretations of and comments about technologies, even without extensive technical expertise, can help develop and validate the democratic process of discussion that is necessary to place technologies within cultural values.

The current data cannot account for silent rejection (or acceptance) of the information through not choosing the share online (although sharing volumes could provide hints), and engagement patterns may be a symptom of general preferences for sharing affirmative information (Muchnik et al., 2003). However, there is suggestion that this statement needs to be qualified in this context, with the point 'interpretations of and comments about' revisited, to consider potential encouraging factors.

Chart 8: Tweet character count

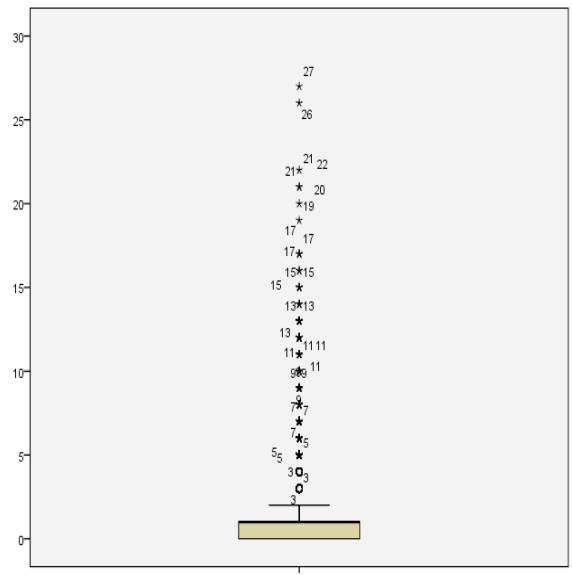


Despite these low engagement levels, four-fifths (81%) of tweets did not utilise the upper Tweet character limit (see chart 8)—although 920 (40%) had a character count of 130 or more, and the median was 120. For comparison, a university of Pennsylvania linguistics professor found that the average word length of Tweets from his university's newspaper was 4.8 characters (Liberman, 2011).

When are people tweeting?

On average, tweets were published by two days since publication —although the longest duration recorded was 27 days, with 353 outliers (three or more days since publication) (see chart 9).

Chart 9: Tweet days since publication



Conversely, that leaves 85% published two days after the article, and 97% within a week.

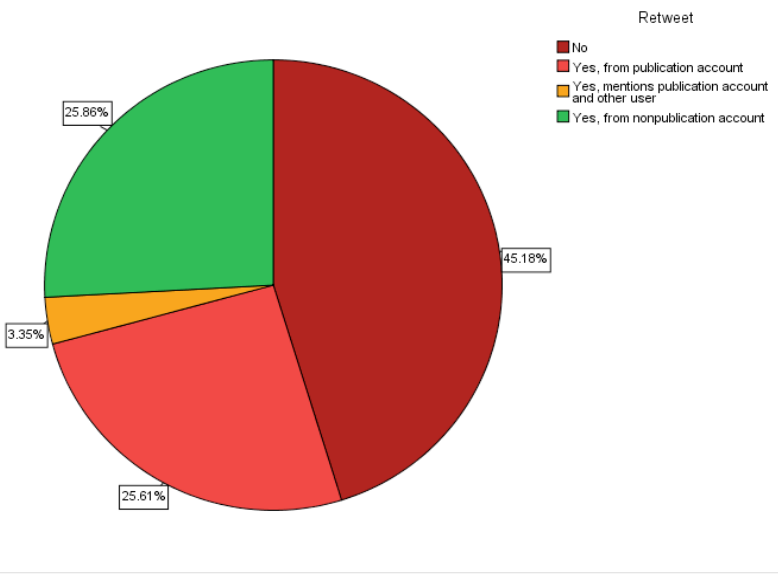
Article visibility on the section homepage may be a contributing factor—albeit with variability deriving from amorphous information travel online (Thorson, 2008). It could also reflect the fleeting nature of information spread on Twitter (You just shared a, 2011).

However, it is unknown whether tweets could spike beyond the measured timeframes. Regardless, the data suggest that that the early part of the article's lifecycle is crucial for social media impact.

From and to where are people tweeting?

About three-quarters of tweets appear to have taken the publication as original source, with only 26% marked as retweets from not mentioning publication accounts, and nearly half (45%) not marked as retweets (see chart 10). Whilst 29% were coded as retweets mentioning the publication account, these may have come from the tweet button on the article page—which, for all publications, renders retweet syntax. (However, the 27% of tweets originated from mobile devices exhibited different patterns).

Chart 10: Tweets marked as retweets



By comparison, a small number of tweets not marked as retweets mentioned other users (which could indicate replies or other interactions)—with 24 mentioning publication accounts only, and 74 mentioning nonpublication accounts. However, due to inconsistent syntax conventions on Twitter, coding could not account for retweets of nonpublication accounts that also mentioned other users. Whilst many of these may be double retweets, it is also possible that retweets contain other forms of interaction. However, a repetition analysis searching for the most-mentioned usernames (including retweets and otherwise) shows that only two of the top 10 are nonpublication accounts, and the top 20 mentioned usernames appeared 919 times altogether—suggesting that engagement tends to centre around certain

users, although a minority of users originated tweets more than once (Hindman, 2009) (see appendix D).

Furthermore, only 4% of tweets were generated from the Twitter reply button, which may be to messages already containing the link, or others that do not. Descriptive statistics of the number of times each of the coded tweets was retweeted, favourite, or replied to paint a similar picture. Table 1 shows that the average number of replies per tweet is much smaller than retweets.

Table 1

Table 1: Types of interactions generated by tweets coded				
Type of interaction generated	Minimum	Maximum	Mean	Standard deviation
Retweets	0	154	6.32	20.056
Favourites	0	20	2.11	6.712
Replies	0	61	.64	1.933

Based on these data, some information is spreading through retweets, which may include responses to or comments aimed at other users. However, tweets containing an explicit indication of direct interaction with other users beyond simply passing on the information were scarce. It thus appears that the publication carries considerable weight as source—with vertical, rather than horizontal information travel prevalent.

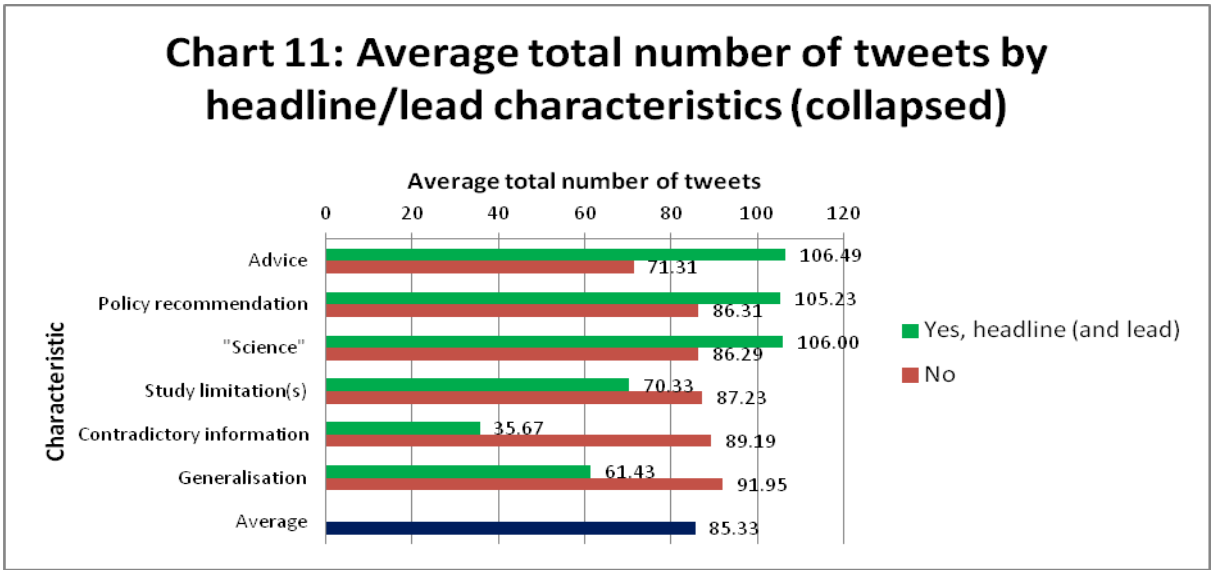
What factors could be contributing to engagement patterns?

Engagement patterns are investigated under the umbrellas of tweet volume and characteristics, respectively—tracing different sources of influence identified in the literature review.

Tweet volume

Whilst significance tests cannot be used on the means of total number of tweets based on article characteristics, descriptive statistics could hint at factors contributing to tweet rate. On average, the articles had one fewer tweet for every extra character in the headline.
Other differences for headline/lead characteristics, in line with previous findings in the literature, include a higher average volume of tweets for articles containing some sort of recommendation—be it on a personal or policy level—in the headline, and lower volumes for

articles indicating some sort of controversy—through highlighting study limitations or contradictions to other knowledge (see chart 11 below). Articles containing the word 'science' in the headline had a higher average tweet volume than those that did not, whilst those containing generalisations had a lower volume.



When considering headline characteristics alongside lead characteristics, however, patterns are inconsistent for all variables apart from generalisations and study limitations—with articles with characteristics present in the lead but not the headline having larger or smaller tweet volumes than both those containing the characteristic in the headline and those containing it in neither the headline nor lead (see chart 12 on page 31). This could indicate that only the headline characteristics are important (although, apart from the 'science' and contradictory information variables, there were no more than two articles per publication containing any of these variables in the lead only). However, many factors that can contribute to tweet rate, which would need to be taken into account when further exploring these differences (for a breakdown by publication, see appendix E).

Tweet characteristics: Summary

Various factors were associated with the presence of headlines or other content, references to article content outside the headline, personal comments, and uncritical assessments. Table 2 (p33) summarises relationship directions (for full model outputs, see appendix F). Only results deemed most important to the analysis are discussed in this section. All results are reported after controlling for all other model variables, with P<=0.05 significance levels.

Tweet characteristics: Scientific influence

The word 'science' was associated with lower engagement levels when considering the presence of headlines or other content. Articles with the word 'science' in the headline had 53% lower odds of only containing content that does not include the headline, and 226% higher odds of including the full headline than otherwise (the difference was also significantly stronger than for the lead). Whilst the variable was not associated with any difference in presence or type of comments, it does appear that it serves as a cue for headline promulgation.

Moreover, tweets linking to articles by science specialists had 27% lower odds of containing personal comments than for articles from generalists (although 73% of sample articles were not written by a science specialist, and only three authors more than five of the articles in the sample—at 14, 9, and 7, respectively).

Notwithstanding limitations to interpretation of article variables, there is some indication that cues indicating scientific authority are associated with reduced levels of overall, if not critical, engagement, and greater emphasis on specific article texts. However, the study source and sex variables were not statistically significant.

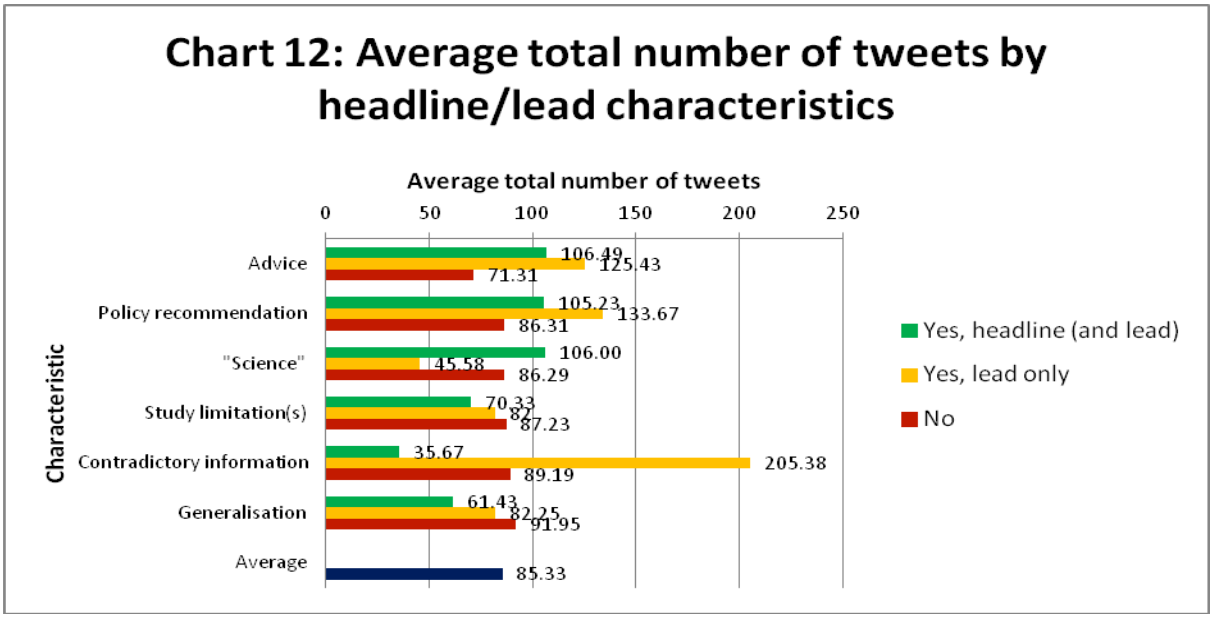


Chart 13 depicts some of the other factors that could be contributing to tweet volume.

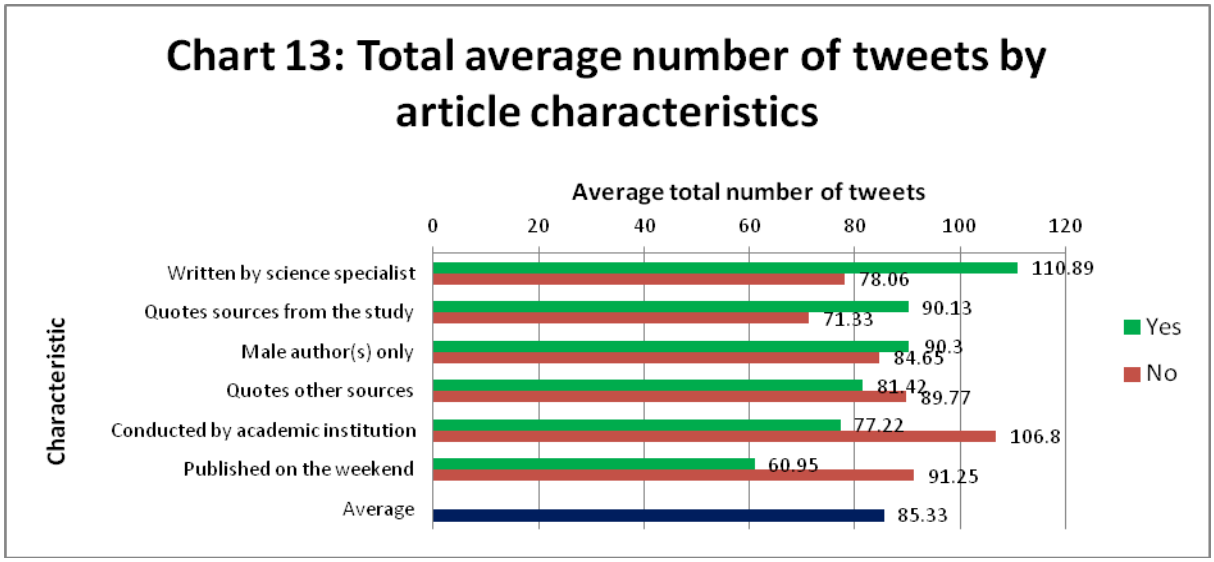


Table 2: Model association direction summary			
Response variable (binary)	Explanatory variable	Base	Association direction
Tweet contains content other than headline	Article generalisation in headline	Binary variable	-
	Retweet mentioning publication account	Nonretweet	-
	Retweet not mentinoing publication account	Nonretweet	+
	Nonretweet mention of other user	Binary variable	+
Tweet contains only content that does not include headline	Article "science" in headline	Binary variable	-
	Article headline character count	Scale variable	+
	Article published on Guardian.co.uk	Published on Dailymail.co.uk	-
	Article published on Telegraph.co.uk	Published on Dailymail.co.uk	-
	Article published on Independent.co.uk	Published on Dailymail.co.uk	-
	Retweet not mentioning publication account	Retweet mentioning publication account	+
	Nonretweet	Retweet mentioning publication account	+
	Nonretweet mention of other user	Binary variable	+
Tweet contains full headline only	Article headline character count	Scale variable	+
	Retweet mentioning publication account	Retweet not mentioning publication account	+
	Nonretweet	Retweet not mentioning publication account	+
	Nonretweet mention of other user	Binary variable	-
Tweet contains headline or headline excerpt only	Article generalisation in headline	Binary variable	+
	Article headline character count	Scale variable	-
	Retweet mentioning publication account	Retweet not mentioning publication account	+
	Nonretweet	Retweet not mentioning publication account	+
	Nonretweet mention of other user	Binary variable	-
Tweet contains full headline (including with other content)	Article "science" in headline	Binary variable	+
	Article headline character count	Scale variable	+
	Article published on Guardian.co.uk	Published on Dailymail.co.uk	+
	Article published on Telegraph.co.uk	Published on Dailymail.co.uk	+
	Article published on Independent.co.uk	Published on Dailymail.co.uk	+
	Retweet mentioning publication account	Retweet not mentioning publication account	+
	Nonretweet	Retweet not mentioning publication account	+
	Nonretweet mention of other user	Binary variable	-
	Tweet originated from mobile device	Binary variable	-
Tweet refers to article lead	Article headline character count	Scale variable	+
Tweet refers to article content outside headline and lead	Article generalisation in headline	Binary variable	-
	Retweet not mentioning publication account	Binary variable	+
Tweet contains user personal comment	Article headline highlights study limitations	Binary variable	+
	Article written by science specialist	Binary variable	-
	Retweet not mentioning publication account	Retweet mentioning publication account	+
	Nonretweet	Retweet mentioning publication account	+
	Nonretweet mention of other user	Binary variable	+
	Tweet generated from Twitter reply button	Binary variable	+
Tweet indicates user uncritical acceptance of information	Article headline highlights study limitations	Binary variable	-
	Article headline highlights conflict with other information	Binary variable	-
	Retweet not mentioning publication account	Retweet mentioning publication account	-
	Nonretweet	Retweet mentioning publication account	-
	Tweet generated from Twitter reply button	Binary variable	-
Retweet not mentioning publication account	Tweet user ID appears more than once in sample	Binary variable	-
	Tweet originated from mobile device	Binary variable	+
	Tweet character count	Scale variable	-
	Tweet contains full headline (including with other content)	Binary variable	-
Retweet mentioning publication account	Tweet user ID appears more than once in sample	Binary variable	-
	Tweet originated from mobile device	Binary variable	-
	Tweet contains full headline (including with other content)	Binary variable	+
Nonretweet	Tweet user ID appears more than once in sample	Binary variable	+
	Tweet originated from mobile device	Binary variable	-
	Tweet contains full headline (including with other content)	Binary variable	-
Tweet generated from Twitter reply button	Tweet originated from mobile device	Binary variable	+
	Tweet contains full headline (including with other content)	Binary variable	-

Tweet characteristics: mainstream media influence

The main characteristics deemed important in light of the potential influence of mainstream media were generalisations, recommendations, and indication of controversy.

Whilst the presence of generalisations did not make a difference to the presence of any types of comments, generalisations in the headline were more likely to be passed on unmodified. For instance, tweets coming from articles containing generalisations in the headline had 66% greater odds of containing the headline only in some form than otherwise (the difference for generalisations in the lead was not statistically significant). Moreover, tweets linking to articles containing generalisations in the headline had 46% lower odds of containing content that does not include the headline than otherwise (again, no significant difference for the lead). Whilst no associations were found for the promulgation of full headlines (as opposed to excerpts), the data indicate that generalisations could be getting passed on in the form that they appear (or further simplified) without reference to other information (articles containing generalisations in the headline also had 68% lower odds of referring to content outside of the headline and lead than otherwise). No association was found for this variable and any type of personal comments.

Whilst the advice and policy variables yielded no significant associations, this in itself is significant, as it indicates that sharing patterns are in line with overall sample trends.

Conversely, tweets linking to articles indicating controversy were less likely to be uncritical. Those linking to articles with headlines highlighting conflict with other information had 72% lower odds of being uncritical than otherwise (the difference for the lead was not statistically significant). Similarly, articles with headlines highlighting study limitations had 71% lower odds of being uncritical than otherwise (all articles highlighting study limitations in the headline also did so in the lead, and only one article highlighted limitations in the lead only). Tweets linking to articles containing the latter characteristic also had 180% higher odds of containing personal comments than otherwise. Whilst these findings are consistent with literature indicating that critical meanings tend to coincide with the degree of closure encoded into the text (Livingstone, 2003), it is worth noting that there were only 42 tweets each linking to six and three articles, respectively, containing these characteristics in the headline.

Article publication was also associated with tweet content breakdown. Whilst editorial decisions could determine different outcomes for publication variables, these were deemed particularly relevant to the influence of technology.

Tweet characteristics: Technological influence

The data suggest that the publication tweet button's pre-generated content could have a bearing on the ways in which information is spreading on Twitter.

Retweets mentioning publication accounts, for instance, had 63% lower odds of containing content apart from the headline than tweets that were not retweets. As the publication tweet button automatically marks tweets as retweets from the publication account, this could indicate that users are not modifying the automatically-generated content before tweeting it.

Whilst Telegraph.co.uk, Guardian.co.uk, and Independent.co.uk were not significantly different from each other in terms of tweet content breakdown, tweets linking to these publications had 75%, 86%, and 225% higher odds, respectively, of containing the full article headline than those linking to Dailymail.co.uk. This could reflect Dailymail.co.uk's practice of modifying content pre-generated from the tweet button, as opposed to just leaving the headline. It could also be a symptom of the average Dailymail.co.uk headline length—which, at 93.92 characters, is 25.35 more than Independent.co.uk's average at 68.57 (with Guardian.co.uk and Telegraph.co.uk at 58.73 and 54.14, respectively). However, the odds do not reflect the relative headline length pattern and, counterintuitively, no consistent associations were observed between headline character count and tweet content. Longer headlines were actually associated with increased odds of containing content apart from the headline, and containing the full headline in any content mix. These factors tie with observations that users are not using the full character limit.

There was no difference between publications (and very little between other variables) with regards to references to article content beyond the headline. Based on this metric, it cannot be inferred from the current data that the location of the tweet button makes a difference to engagement with different parts of the article.

Whilst tweets coming from mobile devices had 24% lower odds of containing the full headline somewhere than those not coming from mobile devices, there was no significant difference for all other response variables analysed. However, more than half of tweets coming from mobile devices were coded as retweets from nonpublication accounts—compared to only 17% of those not coming from mobile devices (see charts 14 and 15).

Chart 14: Tweets coming from mobile devices marked as retweets

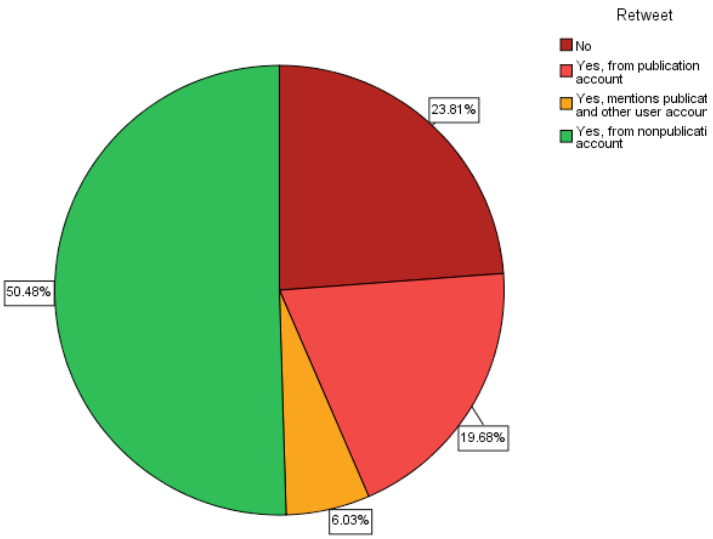
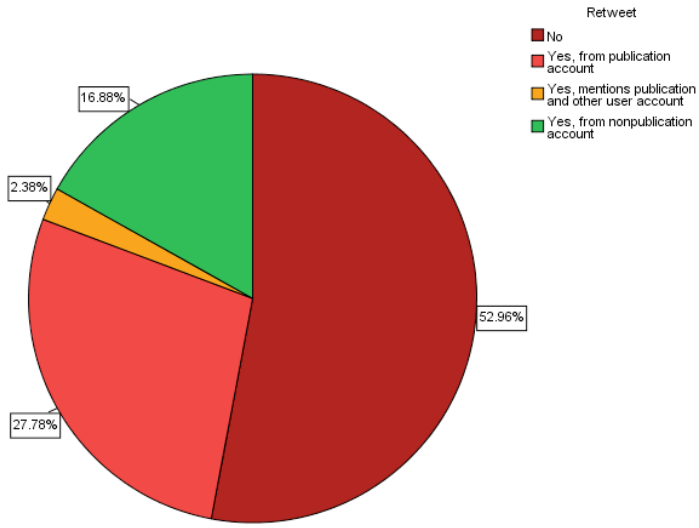


Chart 15: Tweets not coming from mobile devices marked as retweets



Tweets coming from mobile devices had 342% greater odds of being marked as retweets not mentioning publication accounts than otherwise. It thus appears that, despite lack of effect on content per se, information is more likely to spread horizontally on mobile devices, and vertically otherwise. Tweets from mobile devices also had 382% greater odds of being marked as replies. These factors may be due to decreased convenience of visiting publication websites on mobile devices.

Tweet characteristics: network influence

Tweets marked as retweets between users were associated with higher engagement levels, whilst retweets mentioning publication accounts had engagement even than nonretweets. Linking with automatically-generated tweet tweet button content, retweets mentioning publication accounts had 341% higher odds of containing the full headline only than retweets not mentioning publication accounts. Similarly, retweets not mentioning publication accounts had 226%, and nonretweets 143%, higher odds of containing personal comments than retweets mentioning publication accounts. Retweets not mentioning publication

accounts also had 42%, and nonretweets 47%, lower odds of being uncritical than retweets mentioning publication accounts.

It is unknown whether more engaged information is more likely to be passed on between users, or whether users feel more comfortable to engage with information coming from other users than from mainstream media. However, the data do provide clues to the potential contradiction to the herding effect around affirmative messages (Muchnik *et al.,* 2013), and generally increased acceptance of messages coming from peers (Sundar and Nass, 2001). It could also be that peers are participating in filtering information. Whilst it is not possible to glean from the data whether retweets mentioning the publication came from the publication tweet button or the publication Twitter account, it is possible that people are not taking the opportunity to modify the pre-generated content before publishing, or that people are less comfortable modifying the content coming from mainstream media sources. Based on the data on how people are making use of the tweet character limit, it is possible that, more than the technical constraints of the technology, user engagement could be restricted by the speed of information travel online—which is partially supported by the finding from the data that each extra tweet character count is associated with 3% lower odds of the tweet being a retweet not mentioning a publication account (despite higher levels of overall engagement for this variable). In terms of horizontal information flow influence, users that appeared more than once in the sample were less likely to retweet in any form.

Tweets marked as replies or those mentioning other users were also associated with higher levels of engagement but these tweets were in the minority.

What are the implications of these engagement patterns?

The data indicate that a number of factors could be interacting with the audience to shape how scientific information spreads online (Eliasoph, 2004)—with the audience more often than not serving as endorsements for the publications (Thorson, 2008). It is reassuring that information spreading amongst users tends to be more engaged, and tweets coming from mobile devices are more likely to be retweets from other users than otherwise—especially as mobile access to social networking sites is now at 39% of the population and rising (Ofcom, 2013; Brenner and Smith, 2013). However, it appears that more active users are more engaged with publications than other content and, overall, most information flow appears to be vertical. Moreover, in terms of the extent of public engagement, the audience's participatory influence does not appear as strong as technological and active audience enthusiasts would indicate.

Neither the technologies nor the public should be treated as heroes (Eliasoph, 2004). Morley (1993, p14) counters sweeping assumptions about active audiences by warning of the risk '…to underestimate the force of textual determinacy in the construction of meaning from media products, and not only to romanticise improperly the role of the reader, but to risk falling into a complacent relativism, by which the interpretive contribution of the audience is perceived to be of such a scale and range as the render the very idea of media power naïve.' The data suggest that in this context, for better or worse, the democratisation of influence is not as extensive as Gillmor (2009) would suggest. As Hindman (2009) suggests, commercial websites can still play important roles as filters. Through the article tweet button, mainstream media are now even able to put words in people's mouths. Moreover, despite increased access to the public through social media, only one tweet was found coming a study author ('@DailyMailUK Your story about my research is inaccurate. Please remove it. http://t.co/eKARPzqNxu'). Whilst Keen's (2008) concerns about the belittlement of expertise does not appear to stand up in this context (apart from, perhaps, media belittlement of scientific expertise), his concerns about superficial observations rather than considered judgment may not be unfounded.

The technology has already been adopted by the users—who are dealing with online flows of information as they see appropriate—and is unlikely to change unless people vote with their feet. It is thus necessary to consider how to improve outcomes in other ways.

Firstly, headlines may have to be treated as texts in their own right. It is also interesting to note that the article variables associated with higher levels of engagement—particularly those highlighting limitations and conflict with other knowledge early in the article—were scarce in the sample.

This could have implications for what gets promulgated on social media—particularly in light of the overall patterns of engagement, and when considering the presence of generalisations and advice/policy recommendations (see appendix G for descriptive statistics on these variables). Information on the NHS Behind the Headlines service indicates that publications do get it wrong—including in the headline or lead. Whilst this service does not provide comprehensive coverage of assessments on all science-related articles, out of the 19 sample articles that were covered by the service, only 18 did not contain inaccurate or misleading in either the headline or lead information (although this could be due to selection bias), and 15 contained inaccurate or misleading information in the headline—based on information provided by the service (these are marked in appendix B). That said, only four of those

contained inaccurate information: ‘Chips and white bread trigger cravings in the brain’ on Telegraph.co.uk (Brain Scans Link High, 2013), ‘Casual sex makes you depressed and anxious, finds new study’ on Dailymail.co.uk (Casual Sex Linked To, 2013), ‘Choral singing regulates heartbeat’ on Telegraph.co.uk (No Proof that Singing, 2013), and ‘It's NOT an old wives' tale: Cranberry juice really does prevent bladder infections’ on Dailymail.co.uk (Cranberries Not a Proven, 2013), whilst the rest were classified as misleading.

Nonetheless, it was considered that all four of these headlines could be construed as advice by the reader. Depending on offline reception, these factors could all have a bearing on considerations of lifestyle decisions (Leveson, 2012)—bearing in mind that articles containing advice in headlines, relatively, had the highest average total number of tweets. Moreover, the headline proclaiming that casual sex makes people depressed and anxious (despite the lead acknowledging that the direction of causality is unclear) could have implications for perceptions about people with mental health—laced with socially conservative moral undertones (although only 80 tweets linked to this article).

Other headlines combining generalisations with misleading information include: ‘The ‘ladette culture’ is being blamed for an alarming rise in the numbers of women in their thirties and forties dying from alcohol-related conditions’ on Dailymail.co.uk (Media Blames ‘Ladettes’ as, 2013), ‘Prozac Nation: Use of antidepressants in the UK has soared by 500% in the past 20 years’ on Dailymail.co.uk (Prozac Nation’ Claim as, 2013), ‘Working night shifts hits women's fertility: Odd hours 'make it 80% harder to become pregnant'‘ on Dailymail.co.uk, and ‘Working shifts can harm women’s fertility’ on Telegraph.co.uk (Long-term Night Shifts, 2013). The last two could have implications for views on women in the workplace. Both omit that differences were found only after 30 years, by which point a woman would be aged 46, even if she started at 16. Furthermore, whilst Dailymail.co.uk does mention that other lifestyle factors may be contributing to the infertility, it only makes this clear in the final line with the statement: ‘…it is too early to advise stopping shift work. Dr Lavery said: ‘There’s no evidence at the moment that changing your job improves your reproductive outcome.” (although only 28% of the tweets coded for these articles contained only the headline).

Whilst research suggests that, in emergencies, tweet patterns can weed out unfounded rumours (Mendoza *et al.,* 2010), Morris et al. (2012) found that Twitter users did not find accurate information more credible, and Twitter users generally considered science tweets to be more credible than those on politics and entertainment.

Whilst these article characteristic variables did not show different patterns for uncritical tweets, the potential impact of their sharing patterns could be mitigated by letting the readers know early on if there are issues with veracity of the study, in addition to ensuring upmost accuracy. Conversely, false balance has the potential to inspire unfounded critical engagement in the public domain.

This presents a conundrum for audience engagement. Firstly, there is a risk of giving undue weight to certain types of information (Climate: Public Understanding and, 2013)—although this appears to be an issue only in certain cases (to illustrate, 65% of the sample articles did not quote any sources not affiliated with the study, compared to only 17% for study sources). Secondly, when looking at sharing volumes and engagement levels, the patterns found appear to be inverse in most instances; tweets linking to articles with characteristics associated with higher participation volumes had fewer higher-engagement characteristics, and vice-versa. Whilst there is an imperative for exposure diversity (Napoli, 1999) in terms of volume—to facilitate greater public awareness and understanding of science—exposure diversity in terms of the type of engagement may be more important for consumption, as Hall (2003, p. 53) argues: 'Before this message can have an 'effect' (however defined), satisfy a 'need' or be put to 'use', it must first be appropriated as a meaningful discourse and be meaningfully decoded. It is this set of decoded meanings which 'have an effect; influence, entertain, instructor persuade, with very complex perceptual, cognitive, emotional, ideological or behavioural consequences.' This relates back to Napoli's (1999) point that, just because people have access to information doesn't mean that they actually use it. This could also help audiences make more informed decisions about information presented.

As such, quality should be favoured over quantity—to promote the spread of science in a manner that allows audiences to engage with, and thus internalise, accurate information. Leveson's (2012) emphasis on improving accuracy is not an easy task—especially in light of the social media emphasis on headlines, as simplifying science often makes it wrong (Climate: Public Understanding and, 2013) (although no association was found between readability and engagement levels).

The data suggest that the right, and the wrong, types of information have considerable potential to spread on social networks verbatim—especially through headlines and pre-generated publication tweet content. The factors that could mitigate misinformation are also associated with greater audience engagement. Whilst the sample could not provide sufficient information to gauge the difference in importance between headlines and leads for highlighting controversial information, other variables—as well as the overall sharing

patterns—suggest that current approaches to regulating headlines content need to be revisited, including not considering them in isolation, and resolving complaints promptly. It is also necessary to consider other editorial and technological tactics to promote audience engagement. This could, for instance, include modifying the pre-generated content from the publication tweet button in such a way that encourages users to peruse other parts of the articles, seek out the source of the article, or make personal comments on the texts.

CONCLUSION

This paper has situated the current debates surrounding science in the media within the context of social media sharing. The findings of this study suggest that the vertical travel of information from mainstream media remain influential in the audience self-communication process (Castells, 2009). The data indicate that a participatory audience is not necessarily active in meaning-making or critical engagement—at least in terms of information published in the public domain. More often than not, audiences—like the mainstream media before them, serve as little more than a public relations service for the scientific information with which they are presented.

There is a particular emphasis on headlines, which often get passed on verbatim, with little reference to other parts of article, and the limited number of personal comments indicating uncritical acceptance of all content. Articles containing advice in the headlines have higher share volumes, whilst headlines containing generalisations that could feed into stereotypes are more likely to be shared without other content. As mainstream media headlines still leave much to be desired in terms of accuracy—especially as information travels through economic, political, and ideological lenses—this could have implications for the public understanding of science and life-worlds.

Audiences are not necessarily at fault, as there is only so much that can be expected with information flooding in from multiple sources. Thus, it is necessary to consider how best to deal with current participation patterns. These suggest that future regulators should place greater emphasis on the content found in headlines, in light of the potential to for republication on social media, outside of the context of the article in question. They further suggest that the content of the social media sharing button, which to date has escaped attention, is a matter for consideration in terms of content accuracy and the promotion of engagement, which may be influenced by various factors relating to scientific authority,

mainstream media, technology, and the network—particularly in terms of the factors surrounding the still-prevalent vertical travel of information.

ACKNOWLEDGMENTS

Thank you to my supervisor Professor Robin Mansell for her patience. And to other LSE lecturers for chipping in—particularly Dr Alison Powell, Dr Ellen Helsper, and Dr Nick Anstead. Special thanks to my LSE colleagues who also made this dissertation possible: Christian Ledwell for his time, Paul Moura for the sanctuary, and many others for invaluable inspiration throughout the year.

REFERENCES

Adorno, T., and Horkheimer, M. (1944) *The Culture Industry: Enlightenment as Mass Deception in Dialectic of Enlightenment*. New York: Seabury Press.

Agresti, A., and Finlay, B. (2009) *Statistical Methods for the Social Sciences* (4th ed.) Pearson Prentice Hall.

Alberdingk Thijm, Y., Heimans, J., Hurwitz, B., Jiménez, C., and Roth, K. (2012) *First Do No Harm: Changing the Human Rights Landscape*. Retrieved from http://www.fordfoundation.org/issues/freedom-of-expression/advancing-media-rights-and-access/news?id=677

Arnstein, S. (1969) A Ladder of Citizen Participation. *Journal of the American Planning Association*, *35*(4), 216–224.

Bates, J. (2012) This is what modern deregulation looks like: Co-optation and contestation in the shaping of the UK's Open Government Data Initiative. *Journal of Community Informatics*, *8*(2)

Bauer, M. (2000) Chapter 8: Content Analysis. In M. Bauer and G. Gaskell (Eds.), *Qualitative Researching with Text, Image and Sound. A practical handbook for social research*. London: Sage.

Beales, I. (2012) The Editor's Codebook. Press Complaints Commission. Retrieved from http://www.editorscode.org.uk/downloads/codebook/codebook.pdf

Beck, U. (1992) *Risk Society: Towards a New Modernity*. London: SAGE Publications Ltd.
Behind the Headlines. (2013) *NHS choices*. Retrieved from http://www.nhs.uk/news/Pages/NewsIndex.aspx

Berger, J. (2013) *Contagious: Why Things Catch On*. UK: Simon and Schuster.

Bergus, G., Chapman, G., Levy, B., Ely, J., and Oppliger, R. (1998) Clinical diagnosis and the order of information. *PubMed*, *18*(4), 412–7.

Bird, S. E. (2003) Beyond the Audience: Living in a media world. In *The audience in everyday life*. New York: Routledge.

Boczkowski, P. J. (2004) The Mutual Shaping of Technology and Society in Videotex Newspapers: Beyond the Diffusion and Social Shaping Perspectives. *The Information Society*, *20*, 255–267.

Borgatti, S. P., and Halgin, D. S. (2011) On Network Theory. *Organization Science*. doi:10.1287/orsc.1100.0641

Boyd, D., Golder, S., and Lotan, G. (n.d.) Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter DRAFT

Brain scans link high GI diet to food cravings. (2013, June 27) *Behind the Headlines*. Retrieved from http://www.nhs.uk/news/2013/06June/Pages/brain-scans-link-high-GI-diet-to-food-cravings.aspx

Brenner, J., and Smith, A. (2013) *72% of Online Adults are Social Networking Site Users*. US: Pew Research Center's Internet and American Life Project. Retrieved from

http://www.pewinternet.org/~/media//Files/Reports/2013/PIP_Social_networking_sites_update.pdf

Caldwell, C. (2000) Why Do People Join Local Exchange Trading Systems? *International Journal of Community Currency Research*, *4*. Retrieved from http://www.ijccr.group.shef.ac.uk/vol/vol4/Why_Do_People_Join_Local.htm

Carlson, N. (2010, April 14) Twitter Search Is Bigger Than Bing And Yahoo Search Combined. *Www.businessinsider.com*. Retrieved from http://www.businessinsider.com/twitter-search-is-bigger-than-bing-and-yahoo-search-combined-2010-4

Carvalho, A. (2007) Ideological cultures and media discourses on scientific knowledge: re-reading news on

climate change. *Public Understanding of Science*, *16*, 223–243. doi:DOI: 10.1177/0963662506066775

Castells, M. (2009) *Communication Power*. Oxford, UK: Oxford University Press.

Castillo, C., Mendoza, M., and Poblete, B. (2011) Information Credibility on Twitter. Presented at the WWW 2011 – Session: Information Credibility, Hyderabad, India. Retrieved from http://www.ra.ethz.ch/cdstore/www2011/proceedings/p675.pdf

Casual sex linked to depression and anxiety. (2013, July 4) *Behind the Headlines*. NHS choices. Retrieved from http://www.nhs.uk/news/2013/07July/Pages/Casual-sex-linked-to-depression-and-anxiety.aspx

Chan, S. (2012) *Big Data and the Media*. Retrieved from http://www.fordfoundation.org/issues/freedom-of-expression/advancing-media-rights-and-access/news?id=677

Cheerleader or watchdog? (2009) *Nature*, *459*. Retrieved from http://www.nature.com/nature/journal/v459/n7250/full/4591033a.html

Chen, J., Nairn, R., Nelson, L., and Bernstein, M. (2010) Short and tweet: experiments on recommending content from information streams. In H. Chi (Ed.), (pp. 1185–1194) Presented at the SIGCHI Conference on Human Factors in Computing Systems, New York, NY, USA.

Climate: Public understanding and policy implications. (2013) *Parliament*. Retrieved from http://www.parliament.uk/business/committees/committees-a-z/commons-select/science-and-technology-committee/news/130228-climate-change-new-inquiry/

"Climategate." (2009, December 22) *FactCheck.org*. Retrieved from http://www.factcheck.org/2009/12/climategate/
Condit, C. (2004) Science reporting to the public: Does the message get twisted? *Canadian Medical Association Journal*, *170*(9), 1415–1416.

Condit, C. M. (1999) How the public understands genetics: non-deterministic and non-discriminatory interpretations of the "blueprint" metaphor. *Public Understanding of Science*, *8*, 169–180. doi:DOI: 10.1088/0963-6625/8/3/302

Cooper, B. E. J., Lee, W. E., Goldacre, B. M., and Sanders, T. A. B. (2011) The quality of the evidence for dietary advice given in UK national newspapers. *Public Understanding of Science*, 1–10.

Couldry, N. (2001) The Hidden Injuries of Media Power. *Journal of Consumer Culture*, *1*, 155–177.

Cranberries not a proven cure for bladder infection. (2013, July 17) *Behind the Headlines*. NHS choices. Retrieved from http://www.nhs.uk/news/2013/07July/Pages/Cranberries-not-a-proven-cure-for-bladder-infection.aspx

Curran, J. (2002) *Media and Power*. London: Routledge.

De Swert, K. (2012) Calculating inter-coder reliability in media content analysis using Krippendorff's Alpha. University of Amsterdam. Retrieved from http://www.polcomm.org/wp-content/uploads/ICR01022012.pdf

Dearing, J. W. (1995) Newspaper coverage of maverick science: creating controversy through balancing. *Public Understanding of Science*, *4*, 341–361. doi:DOI: 10.1088/0963-6625/4/4/002

Department of Business, Innovation, and Skills. (2011) *Public Attitudes to Science*. UK. Retrieved from http://www.ipsos-mori.com/Assets/Docs/Polls/sri-pas-2011-main-report.pdf

Dutton, W. H., and Jeffreys, P. W. (Eds.) (2010) *World Wide Research: Reshaping the Sciences and Humanities*. Massachusets Institute of Technology.

Eliasoph, N. (2004) Can We Theorize the Press Without Theorizing the Public? *Political Communication*, *21*(3), 297–303.

Ellerbrok, A. (2010) Empowerment: Analyzing Technologies of Multiple Variable Visibility. *Surveillance and Society*, *8*(2), 200–220.

Epstein, S. (1996) *Impure Science: Aids, Activism, and the Politics of Knowledge*. San Diego: University of California Press.

Fierke, K. (2004) World or Worlds? The Analysis of Content and Discourse. *Newsletter of the American Political Science Association Organized Section on Qualitative Methods*, *2*(1), 36–38.

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. doi:10.1177/1461444807075015

Foucault, M. (1978) *The History of Sexuality: 1. The Will to Knowledge*. London: Random House.

Fox, F. (2010, January) Science and the Media: Securing the Future. Science and the Media Expert Group. Retrieved from http://scienceandsociety.bis.gov.uk/media/files/2010/01/Science-and-the-Media-Securing-the-Future.pdf

Fox, S. (2006) *Online Health Search 2006* (PEW INTERNET and AMERICAN LIFE PROJECT) Pew. Retrieved from http://www.pewinternet.org/~/media/Files/Reports/2006/PIP_Online_Health_2006.pdf.pdf

Freelon, D. (2013) ReCal OIR: Ordinal, Interval, and Ratio Intercoder Reliability as a Web Service. *International Journal of Internet Science*, *8*(1), 10–16.

Freund, E. (2010) "*Discuss This Article!" Participatory Uses of Comment Sections on SPIEGEL ONLINE: A Content Analysis* (MSc Politics and Communication) London School of Economics and Political Science, London. Retrieved from http://www.lse.ac.uk/media@lse/research/mediaWorkingPapers/MScDissertationSeries/2010/Freund.pdf

Galloway, A., and Thacker, E. (2007) We're Tired of Trees. In *The Exploit: A Theory of Networks* (Vol. 21) Minneapolis: U Minnesota P.

Gergen, K. J., and Gergen, M. M. (2008) Social Constructionism. *The Sage Encyclopedia of Qualitative Research Methods*. doi:10.4135/9781412963909

Gillespie, G. (2010) The politics of "platforms." *New Media and Society*, *12*. doi:DOI: 10.1177/1461444809342738

Gillmor, D. (2009) *We the Media: Grassroots Journalism By the People, For the People*. Boston, MA: O'Reilly Media.

Goldacre, B. (2008, August 30) The media's MMR hoax. *Bad Science*. Retrieved from http://www.badscience.net/2008/08/the-medias-mmr-hoax/

Google Translate. (2013) Retrieved from http://translate.google.com/

Grilli, R., Ramsay, C., and Minozzi, S. (2009) Mass media interventions: effects on health services utilisation (Review) The Cochrane Collaboration. Retrieved from http://www.cdbph.org/documents/Mass%20media%20interventions%20%20effects%20on%20health%20services%20utilization.pdf

Gross, A. G. (1994) The roles of rhetoric in the public understanding of science. *Public Understanding of Science*, *3*, 3–23.

Gruber, T. (2008) Collective knowledge systems: Where the Social Web meets the Semantic Web. *Web Semantics: Science, Services and Agents on the World Wide Web*, *6*(1), 4–13.

Gunning, R. (1969) The Fog Index After Twenty Years. *Journal of Business Communication*, *6*(2), 3–13.

Hall, S. (1997) The spectacle of the other. In S Hall (Ed.), *Representation: Cultural Representations and Signifying Practices* (pp. 223–290) Thousand Oaks CA: Sage Publications.

Hall, S. (2001) Chapter 7: Foucault: Power, Knowledge and Discourse. In M. Wetherell, S. Taylor, and S. Y. Yates (Eds.), *Discourse theory and practice: a reader* (pp. 72–81) London: Sage.

Hall, Stuart. (2003) Edoding/Decoding. In K. Ross and V. Nightingale (Eds.), *Critical readings media and audiences* (pp. 51–64) Maidenhead: Open University Press.

Han, C. (2012) South African Perspectives on Mobile Phones: Challenging the Optimistic Narrative of Mobiles for Development. *International Journal of Communication*, *6*, 2057–2081.

Hansen, A. (1994) Journalistic practices and science reporting in the British press. *Public Understanding of Science*, *3*, 111–134. DOI: 10.1088/0963-6625/3/2/001

Hawksey, M. (2013) *Twitter Archiving Google Spreadsheet TAGS v5*. Retrieved from http://mashe.hawksey.info/2013/02/twitter-archive-tagsv5/

Helbing, D., and Balietti, S. (2011) From social data mining to forecasting socio-economic crises. *EUROPEAN PHYSICAL JOURNAL-SPECIAL TOPICS*, *195*(1), 3–68. doi:10.1140/epjst/e2011-01401-8
Hester, J. B., and Dougall, E. (2007) The Efficiency of Constructed Week Sampling for Content Analysis of Online News. *Journalism and Mass Communication Quarterly*, *84*(4), 811.824.

Hidalgo, C., A. (2012) *Bringing a New Lens to Old Data*. Retrieved from http://www.fordfoundation.org/issues/freedom-of-expression/advancing-media-rights-and-access/news?id=677

Hindman, M. (2009) *The Myth of Digital Democracy*. Princeton, US: Princeton University Press.

Hopf, T. (2004) Discourse and Content Analysis: Some Fundamental Incompatibilities. *Newsletter of the American Political Science Association Organized Section on Qualitative Methods*, *2*(1), 31–33.

Horrigan, J. B. (2006, March 22) Online News For many home broadband users, the internet is a primary news source. Pew Internet Pew American Life Project. Retrieved from http://www.pewinternet.org/Reports/2006/Online-News-For-many-home-broadband-users-the-internet-is-a-primary-news-source.aspx

How are UK national newspaper sites doing with social media? (2011, November) Retrieved from http://www.cloudninepr.com/roi/tracking-news-sites-success-on-social-media

Java, A., Song, X., Finin, T., and Tseng, B. (2007) Why We Twitter: Understanding Microblogging Usage and Communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis* (pp. 56–65) New York, NY, USA. Retrieved from http://aisl.umbc.edu/resources/369.pdf

Jenkins, H., Purushotma, R., Weigel, M., Clinton, K., and Robison, A. J. (2009) *Confronting the Challenges of Participatory Culture*. Cambridge, Massachusetts and London, England: Massachusetts Institute of Technology.

Jha, A. (2013, August 26) Research finds "US effect" exaggerates results in human behaviour studies. *Telegraph.co.uk*. Retrieved from http://www.theguardian.com/science/2013/aug/26/research-funding-exaggerates-results

Johnson, J. and Joslyn, R. (2001) Document Analysis: Using the written record. In J.

Johnson, J. and Joslyn, R. (Eds.), *Political science research methods* (pp. .237–270) Washington D.C.: CQ Press.

Johnson, T. J., and Kay, B. K. (2004) WAG THE BLOG: HOW RELIANCE ON TRADITIONAL MEDIA AND THE INTERNET INFLUENCE CREDIBILITY PERCEPTIONS OT WEBLOGS AMONG BLOG USERS. *JandMC Quarterly*, *81*(3), 622–642.

Jong, G. W., Stricker, B. H. C., and Sturkenboom, M. C. J. M. (2004) Marketing in the lay media and prescriptions

of terbinafine in primary care: Dutch cohort study. *British Medical Journal*, *328*, 931.

Jordan, D., Lee, R., Ball, F., Donald, R., Montford, A., and Painter, J. Climate: Public understanding and policy implications: Uncorrected transcript of oral evidence. To be published as HC 254-iii. HOUSE OF COMMONS, LONDON: UK PARLIAMENT. Retrieved from http://www.publications.parliament.uk/pa/cm201314/cmselect/cmsctech/uc254-iii/uc25401.htm

Karr, R. (2012) *Big Data and the Just City*. Retrieved from http://www.fordfoundation.org/issues/freedom-of-expression/advancing-media-rights-and-access/news?id=677

Kearney, R. (1994) *Modern Movements in European Philosophy: Phenomenology, Critical Theory, Structuralism* (2nd ed.) Manchester: Manchester University Press.

Keen, A. (2008) *The Cult of the Amateur: How blogs, MySpace, YouTube and the rest of today's user-generated media are killing our culture and economy*. New York: Doubleday.

Kim, B., and Rudin, C. (2013) Learning About Meetings. MIT. Retrieved from http://arxiv.org/pdf/1306.1927v1.pdf

Kim, I., and Kuljis, J. (2010) Applying Content Analysis to Web-based Content. *Journal of Computing and Information Technology*, *18*(4), 369–375. doi:doi:10.2498 /cit.1001924

Kirkpatrick, D., and Smolan, R. (2012) *The Human Face of Big Data*. Retrieved from http://www.fordfoundation.org/issues/freedom-of-expression/advancing-media-rights-and-access/news?id=677

Klein, H. K., and Kleinman, D. L. (2002) The Social Construction of Technology: Structural Considerations. *Science, Technology, and Human Values*, *27*(1), 28–52.

Koepsel, K., Carey, L., and Ron, W. (2010, June 11) Is Social Media "Harvesting" or "Scraping" a Crime? *Lexis Nexis Communities Lexis Hub*. Retrieved from http://www.lexisnexis.com/community/lexishub/blogs/legaltechnologyandsocialmedia/archive/2010/11/06/is-social-media-quot-harvesting-quot-or-quot-scraping-quot-a-crime.aspx

Krippendorf, K. (2004) Chapter 1: History. In *Content Analysis* (2nd ed., pp. 3–17) London: Sage.

Krippendorff, K. (2013) *Content Analysis: An Introduction to Its Methodology* (3rd ed.) Pennsylvania: SAGE Publications, Inc.

Kuhn, T. S. (1962) *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Lenhart, A., and Madden, M. (2005) *Teen Content Creators and Consumers*. PEW INTERNET and AMERICAN LIFE PROJECT. Retrieved from http://www.pewinternet.org/~/media//Files/Reports/2005/PIP_Teens_Content_Creation.pdf.pdf

Lessig, L. (2006) *Code v2.0* (2.0 ed.) Retrieved from http://codev2.cc/download+remix/Lessig-Codev2.pdf

Letouzé, E. (2012) *Big Data for Development: Challenges and Opportunities*. UN Global Pulse. Retrieved from http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf

Leveson, T. R. H. L. J. (2012) *An inquiry into the culture, practices and ethics of the press: Report Volume 2* (pp. 688–693) London: The Leveson Inquiry. Retrieved from http://www.official-documents.gov.uk/document/hc1213/hc07/0780/0780_ii.pdf

Liberman, M. (2011, October 28) Up in ur internets, shortening all the words. *Language Log*. University of Pennsylvania. Retrieved from http://languagelog.ldc.upenn.edu/nll/?p=3532
Limitations of the Study. (2013) *USC Libraries*. University. Retrieved from http://libguides.usc.edu/content.php?pid=83009andsid=616083

Livingstone, S. (2003) *The changing nature of audiences : from the mass audience to the interactive media user*.

London: LSE Research Online. Retrieved from http://eprints.lse.ac.uk/archive/00000417

Locke, S. (2001) Sociology and the public understanding of science: from rationalization to rhetoric. *British Journal of Sociology*, *52*(1), 1–18.

Lombard, M., Snyder-Duch, J., and Campanella Bracken, C. (2002) Content Analysis in Mass Communication Assessment and Reporting of Intercoder Reliability. *Human Communication Research*, *28*(4), 587–604.

Lombard, M., Snyder-Duch, J., and Campanella Bracken, C. (2004) Practical Resources for Assessing and Reporting Intercoder Reliability in Content Analysis Research Projects. Retrieved from http://ils.indiana.edu/faculty/hrosenba/www/Research/methods/lombard_reliability.pdf

Long-term night shifts can "double" breast cancer risk. (2013, July 2) *Behind the Headlines*. NHS choices. Retrieved from http://www.nhs.uk/news/2013/07July/Pages/Long-term-night-shifts-can-double-breast-cancer-risk.aspx

Luntz, F. I. (2013) *Words That Work: It's Not What You Say, it's What People Hear*. New York: Hyperion. Retrieved from http://www.mohamedrabeea.com/books/book1_10476.pdf

Macskassy, S. A., and Michelson, M. (2011) Why Do People Retweet? Anti-Homophily Wins the Day. Association for the Advancement of Artificial Intelligence. Retrieved from http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2790/3291

Manjoo, F. (2013, June 6) You Won't Finish This Article. *Slate*. Retrieved from http://www.slate.com/articles/technology/technology/2013/06/how_people_read_online_why_you_won_t_finish_this_article.html

Mann, M. (1986) *The Sources of Power* (Vol. I: A history of power from the beginning to A.D. 1760) Cambridge, UK: Cambridge University Press. Retrieved from http://www.sscnet.ucla.edu/soc/faculty/mann/Doc5.pdf

Mansell, R. (2012) *Imagining the Internet: Communication, Innovation, and Governance*. Oxford: Oxford University Press.

Mareck, M. (2011, February 22) Privacy and ethics in social media research. *RW CONNECT*. Retrieved from http://rwconnect.esomar.org/2011/02/22/privacy-and-ethics-in-social-media-research/

Matthes, J. (2009) WHAT"S IN A FRAME? A CONTENT ANALYSIS OF MEDIA FRAMING STUDIES IN THE WORLD"S LEADING COMMUNICATIONS JOURNALS, 1990-2005. *Journalism and Mass Communication Quarterly*,
86(349) Retrieved from http://jmq.sagepub.com/content/86/2/349.full.pdf

McLuhan, M., and Fiore, Q. (1967) *The Medium Is the Massage*. New York: John Wiley.

Media blames "ladettes" as alcohol-related deaths rise. (2013, July 19) *Behind the Headlines*. NHS choices. Retrieved from http://www.nhs.uk/news/2013/07July/Pages/media-blames-ladettes-as-alcohol-related-deaths-rise.aspx

Mendoza, M., Poblete, B., and Castillo, C. (2010) Twitter Under Crisis: Can we trust what we RT? Presented at the 1st Workshop on Social Media Analytics (SOMA '10), Washington, DC, USA. Retrieved from http://snap.stanford.edu/soma2010/papers/soma2010_11.pdf

Mischaud, E. (2007) *Twitter: Expressions of the Whole Self An investigation into user appropriation of a web-based communications platform* (MSc in Politics and Communication) London School of Economics and Political Science, London. Retrieved from http://www.lse.ac.uk/media@lse/research/mediaWorkingPapers/MScDissertationSeries/Past/Mishaud_Final.pdf

Moody, J. (2010) Network Structure and Diffusion (Working Paper) Duke university. Retrieved from http://papers.ccpr.ucla.edu/papers/PWP-DUKE-2009-004/PWP-DUKE-2009-004.pdf

Morley, D. (1981) The Nationwide Audience: a critical postscript. In *Television, Audiences and Cultural Studies* (pp. 111–122) Routledge.

Morley, D. (1993) Active audience theory: pendulums and pitfalls. *Journal of Communication*, *43*(4), 13–19.

Morris, M. R., Counts, S., Roseway, A., Hoff, A., and Schwarz, J. (2012) Tweeting is Believing? Understanding Microblog Credibility Perceptions. In *CSCW '12 Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work* (pp. 441–450) New York, NY, USA. Retrieved from http://delivery.acm.org/10.1145/2150000/2145274/p441-morris.pdf?ip=158.143.107.81andacc=ACTIVE%20SERVICEandkey=C2716FEBFA981EF17A3E66E77AF11B8D1C5628CCCCDC3C2FandCFID=232452762andCFTOKEN=96460842and__acm__=1373380487_32c443a854644e7fe0f1fa12146c3f09

Mosco, V. (1996) *The Political Economy of Communication: Rethinking and Renewal* (2nd ed.) SAGE Publications Ltd.

Muchnik, L., Aral, S., and Taylor, S. J. (2013) Social Influence Bias: A Randomized Experiment. *Science*, *341*(6146), 647–651. doi:10.1126/science.1240466

Napoli, P. M. (1999) Deconstructing the Diversity Principle. *Journal of Communication*, *49*(4), 7–34. doi:10.1111/j.1460-2466.1999.tb02815.x

Nass, C., and Steuer, J. (1993) Voices, Boxes, and Sources of Messages. *Human Communication Research*, *19*(4), 504–527.

Neuendorf, K. A. (2004) Content Analysis: A Contrast and Complement to Discourse Analysis. *Newsletter of the American Political Science Association Organized Section on Qualitative Methods*, *2*(1), 33–36.

Nielson, J. (2006, October 9) Participation Inequality: Encouraging More Users to Contribute. Nielson Norman Group. Retrieved from http://www.nngroup.com/articles/participation-inequality/

Nisbet, M. C., Brossard, D., and Kroepsch, A. (2003) Framing Science The Stem Cell Controversy in an Age of Press/Politics. *The International Journal of Press/Politics*, *8*(2), 36–70. doi:10.1177/1081180X02251047

No proof that singing in a choir is good for heart. (2013, July 9) *Behind the Headlines*. NHS choices. Retrieved from http://www.nhs.uk/news/2013/07July/Pages/No-proof-that-singing-in-a-choir-is-good-for-heart.aspx

Nuttall, C. (2007, April 10) Micro-bloggers of the world keep it short. *Ft.com*. Retrieved from http://www.ft.com/cms/s/0/9196ee56-e700-11db-9034-000b5df10621.html#axzz2cnbcGM4R

O'Baoill, A. (2000) Slashdot and the Public Sphere. *First Monday*, *5*(9) Retrieved from http://firstmonday.org/issues/issue5_9/baoill/index.html

O'Connor, C., and Rees, G. (2012) Neuroscience in the Public Sphere, *74*(2) Retrieved from http://www.sciencedirect.com/science/article/pii/S0896627312003303

Ofcom. (2013) *Communications Market Report: Scotland*. Ofcom. Retrieved from http://stakeholders.ofcom.org.uk/binaries/research/cmr/cmr13/2013_CMR_Scotland.pdf

Ostertag, S. (2010) Establishing news confidence: a qualitative study of how people use the news media to know the news-world. *Media, Culture and Society*. Retrieved from http://mcs.sagepub.com.gate2.library.lse.ac.uk/content/32/4/597.full.pdf

Parliamentary Inquiry into Science Reporting: Where are the Sceptics? (2013, September 9) *LSE Media Policy Project*.
London School of Economics and Political Science. Retrieved from http://blogs.lse.ac.uk/mediapolicyproject/2013/09/09/parliamentary-inquiry-into-science-reporting-where-are-the-sceptics/

PCC Judgement On Daily Mail HPV Article. (2012, February 27) *Stuff and Nonsense*. Retrieved from http://jdc325.wordpress.com/2012/02/27/pcc-judgement-on-daily-mail-hpv/

Peduzzi, P., Concato, J., Kemper, E., Holford, T., and Feinstein, A. (1996) A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, *49*, 1373–1379.

Petrovic , S., Osborne, M., McCreadie, R., Macdonald, C., Ounis, I., and Shrimpton, L. (2013) Can Twitter replace Newswire for breaking news? Presented at the 7th International AAAI Conference On Weblogs And Social Media, Boston, US. Retrieved from http://demeter.inf.ed.ac.uk/cross/docs/sasaBreakingICWSM.pdf

Pinch, T. J., and Bijker, W. E. (1984) The Social Construction of Facts and Artefacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other. *Social Studies of Science*, *14*(3), 399–441.

Ponsford, D. (2013, May 23) Telegraph website traffic continues to grow despite launch of metered paywall. *Www.pressgazette.co.uk*. Retrieved from http://www.pressgazette.co.uk/adoption-metered-paywall-has-little-effect-telegraphs-overall-website-traffic

Press, A. L. (2006) Audience Research in the Post-Audience Age: An Introduction to Barker and Morley. *The Communication Review*, *9*(2), 93–100.

"Prozac nation" claim as antidepressant use soars. (2013, July 8) *Behind the Headlines*. NHS choices. Retrieved from http://www.nhs.uk/news/2013/07July/Pages/Prozac-nation-claim-as-antidepressant-use-soars.aspx

Purcell, K., Rainie, L., Mitchell, A., Rosenstiel, T., and Olmstead, K. (2010, March 1) Understanding the participatory news consumer How internet and cell phone users have turned news into a social experience. Pew Research Center. Retrieved from http://www.pewinternet.org/Reports/2010/Online-News.aspx

Rabe-Hesketh, S., and Skrondal, A. (Stata Press) *Multilevel and Longitudinal Modelling Using Stata* (2nd ed.) Stata Press.

Rafaeli, S., and Sudweeks, F. (1997) Networked Interactivity. *Journal of Computer-Mediated Communication*, *2*(4) Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.1997.tb00201.x/full

Ramsey, S. (1994) Science and technology: when do they become front page news? *Public Understanding of Science*, *3*(71) doi:10.1088/0963-6625/3/1/005

Ravikumar, S., Balakrishnan, R., and Subbarao, K. (2012) Ranking Tweets Considering Trust and Relevance. Presented at the Ninth International Workshop on Information Integration on the Web, New York, NY, USA. Retrieved from http://dl.acm.org/citation.cfm?id=2331805

Riffe, D., Lacy, S., and Frederick, G. (2005) *Analyzing Media Messages: Using Quantitative Content Analysis in Research* (2nd ed.) Routledge.

Robbins, M. (2010, September 27) This is a news website article about a scientific paper. *The Lay Scientist*. www.guardian.co.uk. Retrieved from http://www.guardian.co.uk/science/the-lay-scientist/2010/sep/24/1

Robbins, M. (2011, June 28) "Science journalists have forgotten how to be journalists." *Guardian.co.uk*. Retrieved from http://www.theguardian.com/science/the-lay-scientist/2011/jun/28/1?guni=Article:in%20body%20link

Robbins, M. (2012a, March 1) Recent PCC complaints highlight Leveson's importance for science. *Guardian.co.uk*. Retrieved from http://www.theguardian.com/science/the-lay-scientist/2012/mar/01/1

Robbins, M. (2012b, November 29) Leveson on science reporting. *The Lay Scientist*. The Guardian. Retrieved from http://www.guardian.co.uk/science/the-lay-scientist/2012/nov/29/leveson-inquiry-science-scepticism

Rosen, J. (2007, June 27) The People Formerly Known as the Audience. Retrieved from http://archive.pressthink.org/2006/06/27/ppl_frmr.html

Rosengren, K. E. (Ed.) (1981) *Advances in Content Analysis*. Beverley Hills, California: Sage.
Said, E. (1978) *Orientalism*. London: Penguin.

Scheufele, D.A. (1999) Framing as a theory of media effects. *Journal of Communication*, *49*(1), 103–122.

Scheufele, Dietram A, and Tewksbury, D. (2006) Framing, Agenda Setting, and Priming: The Evolution of Three Media Effects Models. *Journal of Communication*, *57*(1), 9–20.

Schreier, M. (2012) *Qualitative Content Analysis in Practice*. London: Sage.

Shah, D.V., Jack, M., and Yoon, S-H.Y. (2001) Communication, Context, and Community: An Exploration of Print, Broadcast, and Internet Influences. *Communication Research*, *28*, 464–506.

Shoemaker, P. J., and Rees, S. D. (1996) *MEDIATING THE MESSAGE: Theories of Influences on Mass Media Content* (2nd ed.) USA: Longman Publishers. Retrieved from http://shoemaker.syr.edu/docs/mediating-the-message-2nd-edition-1996-shoemaker-reese.pdf

Shum, S. B., Aberer, K., and Schmidt, A. (2012) Towards a global participatory platform Democratising open data, complexity science and collective intelligence. *EUROPEAN PHYSICAL JOURNAL-SPECIAL TOPICS*, *214*(1), 109–152. doi:10.1140/epjst/e2012-01690-3

Silverstone, R., and Sujon, Z. (2003) Urban Tapestries: Experimental Ethnography, Technological Identities and Place. MEDIA@LSE Electronic Working Papers. Retrieved from http://www2.lse.ac.uk/media@lse/research/mediaWorkingPapers/pdf/EWP07.pdf

Stein, L. (2004) Understanding Speech Rights: Defensive and Empowering Approaches to the First Amendment. *Media, Culture and Society*, *26*, 102–120.

Stereotype. (2013) *Oxford*. Oxford Dictionaries Online. Retrieved from http://oxforddictionaries.com/definition/english/stereotype

Sternthal, B., Phillips, L., and Dholakia, R. (1978) The persuasive effect of source credibility: a situational analysis. *Public Opin Q, Fall; 42*(3), 285–314.

Suh, B., Hong, L., and Pirolli, P. (2010) Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network. In H. Chi (Ed.), (pp. 177–184) Presented at the Social Computing (SocialCom), 2010 IEEE Second International Conference, Palo Alto, CA, U.S.A.

Sundar, S. S., and Nass, C. (2001) Conceptualizing Sources in Online News. *Journal of Communication*, *51*(1), 52–72.

Taleb, N. (2013, February 2) Beware the Big Errors of "Big Data." *Wired*, *Online*, 1. Talk Back to Media. (n.d.)

Tewksbury, D. (2003) What do Americans really want to know? Tracking the behavior of news readers on the internet. *Journal of Communication*, *53*(4), 694.

Text Content Analysis Tool. (2013) *Using English*. Retrieved from http://www.usingenglish.com/resources/text-statistics.php

The people formerly known as the audience. (2011, July 7) *The Economist*. Retrieved from http://www.economist.com/node/18904124

Thompson, J. (1990) Ideology in Modern Societies. In *Ideology and Modern Culture: Social Theory in the Era of Mass Communication*. Cambridge: Polity.

Thorson, E. (2008) CHANGING PATTERNS OF NEWS CONSUMPTION AND PARTICIPATION. *Information, Communication and Society*, *11*(4), 473–489. doi:10.1080/13691180801999027

Topsy Pro Analytics. (2013) Retrieved from http://about.topsy.com/products/proanalytics/overview/

Van der Graaf, A. A. C. G. (2009) *Designing for Mod Development: User creativity as product development strategy on the firm-hosted 3D software platform*. London School of Economics and Political Science, London.

Van Dijck, J. (2008) Users like you? Theorizing agency in user-generated content. *Media, Culture and Society*, *31*, 41–58. doi:DOI: 10.1177/0163443708098245

Wajcman, J., and MacKenzie, D. (Eds.) (1999) Introductory essay: the social shaping of technology. In *The social shaping of technology* (2nd ed.) Buckingham, UK: Open University Press. Retrieved from http://eprints.lse.ac.uk/28638/1/Introductory%20essay%20(LSERO)pdf

Wang, X. (2006) EXPLORING SAMPLE SIZES FOR CONTENT ANALYSIS OF ONLINE NEWS SITES. Association for Education in Journalism and Mass Communication. Retrieved from http://www1.usfsp.edu/journalism/showcase/documents/wangSampleSizesPaper.pdf

Weaver, D. A., and Bimber, B. (2008) Finding News Stories: A Comparison of Searches Using Lexisnexis and Google News. *Journalism and Mass Communication Quarterly*, *85*(3), 515–530.

Wellman, B. (2001) Physical Place and CyberPlace: The Rise of Personalized Networking. *International Journal of Urban and Regional Research*, *25*.

Winner, L. (1986) DO ARTIFACTS HAVE POLITICS? In *The whale and the reactor: a search for limits in an age of high technology* (pp. 19–39) Chicago: University of Chicago Press.

Winner, L. The Societal Implications of Nanotechnology (2003) Washington. Retrieved from http://homepages.rpi.edu/~winner/testimony.htm

Wren-Lewis, J. (1983) The encoding / decoding model: criticisms and redevelopments for research on decoding. *Media, Culture and Society*, 175–197.

You just shared a link. How long will people pay attention? (2011, September 6) *bitly blog*. www.bitly.com. Retrieved from http://blog.bitly.com/post/9887686919/you-just-shared-a-link-how-long-will-people-pay

Zhang, Y., and Wildemuth, B. M. (2009) Qualitative analysis of content. In *Applications of social research methods to questions in information and library science* (pp. 308–319) Westport, Connecticut: Libraries Unlimited.

Zuckerman, E. (Moderator) (2012) *Big Data: Big Challenges and Big Opportunities*. Retrieved from http://www.fordfoundation.org/issues/freedom-of-expression/advancing-media-rights-and-access/news?id=677

APPENDICES

Appendix A: Publication website traffic and social media impact

Table 3: Mainstream media website traffic and social media sharing volumes		
Publication	Total unique website visitors (as of April 2013)	Total links shared per week (as of November 2011)
Dailymail.co.uk	7,833,182	2,908,779
Guardian.co.uk	4,771,866	2,587,258
Telegraph.co.uk	3,041,594	879,783
Indpendent.co.uk	1,131,150	617,148

Appendix B: Repetition analyses of originating and mentioned users

Table 5: Most-mentioned usernames in sample	
Username	Number of mentions
1) ~@guardian	288
2) ~@telegraph	142
3) ~@mailonline	110
4) ~@dailymailus	54
5) ~@guardianscience	51
6) ~@independent	35
7) ~@telegraphnews	34
8) ~@lilt_bologna	31
9) ~@telegraphsci	30
10) ~@unicef	28
11) ~@drudge_report	19
12) ~@draliceroberts	16
13) ~@guardianeco	13
14) ~@soilassociation	13
15) ~@profdavidnutt	12
16) ~@soozaphone	11
17) ~@actionhappiness	10
18) ~@mhf_tweets	8
19) ~@tele_education	7
20) ~@theskinclinic7	7

Table 6: Most frequent tweet originators	
Username	Total number of tweets
kazi_uk	7
kazi_canada	7
naturalmedicin	7
healthjournal	6
talkhealth	6
socialsectoruk	6
oldermenshealth	6
smootherofknots	6
healthierlifeco	5
topsciencenews	5
clientearth	4
guardianscience	4
telegraphsci	4
thewomenslawyer	4
itsmotherswork	4
blvlaw	3
superiorgent	3
collegenatmed	3
adamovichc	3
earlyyearscpd	3
gozde786	3
yesenia25761122	3
conserva_tweet	3
phrcleeds	3
sciencetake	3
myrssfeed1	3
ukmentalhealth	3
newsupdte	3
newsatweb	3
stillsafe	3
csfappliedsci	3
pirwany	3
telegraphnews	3
beanfreaks	3
gettingsome	3
rishumongadoc	3

Appendix C: Tweet volume breakdown by publication

Table 7: Total tweet volume for variables by publication							
		Average tweet volume	Total cases			Average tweet volume	Total cases
Headline/lead characteristics							
Full breakdown				Headline vs rest			
Advice							
Dailymail.co.uk	No	34.27	26	Dailymail.co.uk	No	34.78	28
Dailymail.co.uk	Yes, headline (and lead)	78.35	31	Dailymail.co.uk	Yes	78.35	31
Dailymail.co.uk	Yes, lead only	41.5	2				
Guardian.co.uk	No	175.43	14	Guardian.co.uk	No	158.75	16
Guardian.co.uk	Yes, headline (and lead)	213.6	10	Guardian.co.uk	Yes	213.6	10
Guardian.co.uk	Yes, lead only	42	2				
Telegraph.co.uk	No	41.15	27	Telegraph.co.uk	No	51.76	29
Telegraph.co.uk	Yes, headline (and lead)	106.41	17	Telegraph.co.uk	Yes	106.41	17
Telegraph.co.uk	Yes, lead only	195	2				
Independent.co.uk	No	51.67	3	Independent.co.uk	No	235.26	4
Independent.co.uk	Yes, headline (and lead)	40.67	3	Independent.co.uk	Yes	40.67	3
Independent.co.uk	Yes, lead only	321	1				
Policy							
Dailymail.co.uk	No	58.09	58	Dailymail.co.uk	No	58.09	58
Dailymail.co.uk	Yes, headline (and lead)	34	1	Dailymail.co.uk	Yes	34	1
Dailymail.co.uk	Yes, lead only	0	0				
Guardian.co.uk	No	171.91	23	Guardian.co.uk	No	173.4	25
Guardian.co.uk	Yes, headline (and lead)	341	1	Guardian.co.uk	Yes	341	1
Guardian.co.uk	Yes, lead only	190.5	2				
Telegraph.co.uk	No	75.44	43	Telegraph.co.uk	No	74.18	44
Telegraph.co.uk	Yes, headline (and lead)	23	2	Telegraph.co.uk	Yes	23	2
Telegraph.co.uk	Yes, lead only	20	1				
Independent.co.uk	No	85.43	7	Independent.co.uk	No	85.43	7
Independent.co.uk	Yes, headline (and lead)	0	0	Independent.co.uk	Yes	0	0
Independent.co.uk	Yes, lead only	0	0				
"Science"							

Dailymail.co.uk	No	61.59	51	Dailymail.co.uk	No	58.35	58
Dailymail.co.uk	Yes, headline (and lead)	19	1	Dailymail.co.uk	Yes	19	1
Dailymail.co.uk	Yes, lead only	34.71	7				
Guardian.co.uk	No	180.33	24	Guardian.co.uk	No	180.33	24
Guardian.co.uk	Yes, headline (and lead)	174	2	Guardian.co.uk	Yes	174	2
Guardian.co.uk	Yes, lead only	0	0				
Telegraph.co.uk	No	73.73	40	Telegraph.co.uk	No	72.29	45
Telegraph.co.uk	Yes, headline (and lead)	57	1	Telegraph.co.uk	Yes	57	1
Telegraph.co.uk	Yes, lead only	60.8	5				
Independent.co.uk	No	85.43	7	Independent.co.uk	No	85.43	7
Independent.co.uk	Yes, headline (and lead)	0	0	Independent.co.uk	Yes	0	0
Independent.co.uk	Yes, lead only	0	0				
Study limitations							
Dailymail.co.uk	No	57.26	58	Dailymail.co.uk	No	57.68	59
Dailymail.co.uk	Yes, headline (and lead)	0	0	Dailymail.co.uk	Yes	0	0
Dailymail.co.uk	Yes, lead only	82	1				
Guardian.co.uk	No	186.46	24	Guardian.co.uk	No	186.46	24
Guardian.co.uk	Yes, headline (and lead)	100.5	2	Guardian.co.uk	Yes	100.5	2
Guardian.co.uk	Yes, lead only	0	0				
Telegraph.co.uk	No	73.33	45	Telegraph.co.uk	No	73.33	45
Telegraph.co.uk	Yes, headline (and lead)	10	1	Telegraph.co.uk	Yes	10	1
Telegraph.co.uk	Yes, lead only	0	0				
Independent.co.uk	No	85.43	7	Independent.co.uk	No	85.43	7
Independent.co.uk	Yes, headline (and lead)	0	0	Independent.co.uk	Yes	0	0
Independent.co.uk	Yes, lead only	0	0				
Contradictory information							
Dailymail.co.uk	No	47.22	50	Dailymail.co.uk	No	59.26	54
Dailymail.co.uk	Yes, headline (and lead)	40.6	5	Dailymail.co.uk	Yes	40.6	5
Dailymail.co.uk	Yes, lead only	209.75	4				
Guardian.co.uk	No	175.09	23	Guardian.co.uk	No	179.85	26
Guardian.co.uk	Yes, headline (and lead)	0	0	Guardian.co.uk	Yes	0	0
Guardian.co.uk	Yes, lead only	216.33	3				
Telegraph.co.uk	No	71.45	44	Telegraph.co.uk	No	73.31	45
Telegraph.co.uk	Yes, headline (and lead)	11	1	Telegraph.co.uk	Yes	11	1

Telegraph.co.uk	Yes, lead only	155	1				
Independent.co.uk	No	85.43	7	Independent.co.uk	No	85.43	7
Independent.co.uk	Yes, headline (and lead)	0	0	Independent.co.uk	Yes	0	0
Independent.co.uk	Yes, lead only	0	0				
Generalisation							
Dailymail.co.uk	No	61.42	45	Dailymail.co.uk	No	60.52	46
Dailymail.co.uk	Yes, headline (and lead)	47.62	13	Dailymail.co.uk	Yes	47.62	13
Dailymail.co.uk	Yes, lead only	20	1				
Guardian.co.uk	No	172.57	23	Guardian.co.uk	No	176	24
Guardian.co.uk	Yes, headline (and lead)	226	2	Guardian.co.uk	Yes	226	2
Guardian.co.uk	Yes, lead only	255	1				
Telegraph.co.uk	No	80.24	37	Telegraph.co.uk	No	77.51	39
Telegraph.co.uk	Yes, headline (and lead)	41	7	Telegraph.co.uk	Yes	41	7
Telegraph.co.uk	Yes, lead only	27	2				
Independent.co.uk	No	90.5	6	Independent.co.uk	No	90.5	6
Independent.co.uk	Yes, headline (and lead)	55	1	Independent.co.uk	Yes	55	1
Independent.co.uk	Yes, lead only	0	0				
Physical response variable							
Dailymail.co.uk	No	41.29	35				
Dailymail.co.uk	Yes	81.58	24				
Guardian.co.uk	No	196	11				
Guardian.co.uk	Yes	168	15				
Telegraph.co.uk	No	69.93	28				
Telegraph.co.uk	Yes	75.11	18				
Independent.co.uk	No	44.4	5				
Independent.co.uk	Yes	188	2				
Personal-level action can affect outcome							
Dailymail.co.uk	No	31.61	31				
Dailymail.co.uk	Yes	86.54	28				
Guardian.co.uk	No	195.69	16				
Guardian.co.uk	Yes	154.5	10				
Telegraph.co.uk	No	46.27	26				
Telegraph.co.uk	Yes	105.35	20				
Independent.co.uk	No	106.2	5				
Independent.co.uk	Yes	33.5	2				
Article characteristics							
Written by science specialist							
Dailymail.co.uk	No	59.33	55				
Dailymail.co.uk	Yes	35	4				

Guardian.co.uk	No	159.44	18				
Guardian.co.uk	Yes	225.75	8				
Telegraph.co.uk	No	54.9	21				
Telegraph.co.uk	Yes	86.28	25				
Independent.co.uk	No	85.43	7				
Independent.co.uk	Yes	0	0				
Quotes sources from the study							
Dailymail.co.uk	No	32.17	6				
Dailymail.co.uk	Yes	60.57	53				
Guardian.co.uk	No	116	9				
Guardian.co.uk	Yes	213.65	17				
Telegraph.co.uk	No	47.38	8				
Telegraph.co.uk	Yes	77.13	38				
Independent.co.uk	No	96	1				
Independent.co.uk	Yes	83.67	6				
Male author(s) only							
Dailymail.co.uk	No	62.89	44				
Dailymail.co.uk	Yes	42.4	15				
Guardian.co.uk	No	170.39	18				
Guardian.co.uk	Yes	201.13	8				
Telegraph.co.uk	No	57.05	20				
Telegraph.co.uk	Yes	83.42	26				
Independent.co.uk	No	68	2				
Independent.co.uk	Yes	92.4	5				
Quotes other sources							
Dailymail.co.uk	No	50.59	34				
Dailymail.co.uk	Yes	67.32	25				
Guardian.co.uk	No	178.29	17				
Guardian.co.uk	Yes	182.78	9				
Telegraph.co.uk	No	82.26	34				
Telegraph.co.uk	Yes	42.75	12				
Independent.co.uk	No	106.2	5				
Independent.co.uk	Yes	33.5	2				
Conducted by academic institution							
Dailymail.co.uk	No	78.13	16				
Dailymail.co.uk	Yes	50.07	43				
Guardian.co.uk	No	192.88	16				
Guardian.co.uk	Yes	159	10				
Telegraph.co.uk	No	29	11				
Telegraph.co.uk	Yes	85.46	35				
Independent.co.uk	No	75.5	2				
Independent.co.uk	Yes	89.4	5				

Appendix D: Models output

Tweet contains content other than headline

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	-.353	.4857	-1.305	.599	.528	1	.468
[Article generalisation in headline=.00]	-.620	.2351	-1.081	-.159	6.948	1	.008
[Article generalisation in headline=1.00]	0[a]	.	.	.	.	.	.
[Retweet mentioning publication account=.00]	-1.006	.1416	-1.284	-.729	50.512	1	.000
[Retweet mentioning publication account=1.00]	0[a]	.	.	.	.	.	.
[Retweet not mentioning publication account=.00]	.517	.1332	.256	.778	15.067	1	.000
[Retweet not mentioning publication account=1.00]	0[a]	.	.	.	.	.	.
[Nonretweet mention of other user=.00]	1.020	.4179	.201	1.840	5.962	1	.015
[Nonretweet mention of other user=1.00]	0[a]	.	.	.	.	.	.
(Scale)	1						

Dependent Variable: Tweet contains content other than headline
Model: (Intercept), Article generalisation in headline, Retweet mentioning publication account, Retweet not mentioning publication account, Nonretweet mention of other user
a. Set to zero because this parameter is redundant.

Tweet contains only content that does not include headline

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	.379	.6946	-.983	1.740	.297	1	.585
[Article "science" in headline=.00]	-.753	.2161	-1.176	-.329	12.135	1	.000
[Article "science" in headline=1.00]	0[a]	.	.	.	.	.	.
Article headline character count	.012	.0052	.002	.022	5.119	1	.024
[Article published on Guardian.co.uk=.00]	-.607	.2824	-1.160	-.053	4.618	1	.032
[Article published on Guardian.co.uk=1.00]	0[a]	.	.	.	.	.	.
[Article published on Telegraph.co.uk=.00]	-.676	.2712	-1.207	-.144	6.210	1	.013
[Article published on Telegraph.co.uk=1.00]	0[a]	.	.	.	.	.	.
[Article published on Independent.co.uk=.00]	-1.295	.5663	-2.405	-.185	5.231	1	.022
[Article published on Independent.co.uk=1.00]	0[a]	.	.	.	.	.	.
[Retweet not mentioning publication account=.00]	1.574	.2451	1.093	2.054	41.231	1	.000

[Retweet not mentioning publication account=1.00]	0[a]	.	.	.	.	.	.
[Nonretweet=.00]	1.286	.2381	.819	1.753	29.173	1	.000
[Nonretweet=1.00]	0[a]	.	.	.	.	.	.
[Nonretweet mention of other user=.00]	1.023	.2515	.530	1.516	16.542	1	.000
[Nonretweet mention of other user=1.00]	0[a]	.	.	.	.	.	.
(Scale)	1						

Dependent Variable: Tweet contains only content that does not include headline
Model: (Intercept), Article "science" in headline, Article headline character count, Article published on Guardian.co.uk, Article published on Telegraph.co.uk, Article published on Independent.co.uk, Retweet not mentioning publication account, Nonretweet, Nonretweet mention of other user
a. Set to zero because this parameter is redundant.

Tweet contains full headline only

Parameter Estimates

			95% Wald Confidence Interval		Hypothesis Test		
Parameter	B	Std. Error	Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	-.256	.6811	-1.591	1.079	.141	1	.707
Article headline character count	.014	.0038	.007	.021	13.578	1	.000
[Retweet mentioning publication account=.00]	1.485	.1679	1.155	1.814	78.152	1	.000
[Retweet mentioning publication account=1.00]	0[a]	.	.	.	.	.	.
[Nonretweet=.00]	.432	.1421	.153	.710	9.232	1	.002
[Nonretweet=1.00]	0[a]	.	.	.	.	.	.
[Nonretweet mention of other user=.00]	-1.356	.5719	-2.477	-.235	5.622	1	.018
[Nonretweet mention of other user=1.00]	0[a]	.	.	.	.	.	.
(Scale)	1						

Dependent Variable: Tweet contains full headline only
Model: (Intercept), Article headline character count, Retweet mentioning publication account, Nonretweet, Nonretweet mention of other user
a. Set to zero because this parameter is redundant.

Tweet contains headline or headline excerpt only

Parameter Estimates

			95% Wald Confidence Interval		Hypothesis Test		
Parameter	B	Std. Error	Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	.488	.6097	-.707	1.683	.641	1	.423
[Article generalisation in headline=.00]	.506	.2201	.075	.937	5.287	1	.021
[Article generalisation in headline=1.00]	0[a]	.	.	.	.	.	.
Article headline character count	-.011	.0035	-.018	-.005	10.768	1	.001
[Retweet mentioning publication account=.00]	1.557	.1826	1.199	1.915	72.712	1	.000
[Retweet mentioning publication account=1.00]	0[a]	.	.	.	.	.	.
[Nonretweet=.00]	.467	.1373	.198	.737	11.585	1	.001

[Nonretweet=1.00]	0[a]	.	.	.	.	.	.
[Nonretweet mention of other user=.00]	-1.257	.4633	-2.165	-.349	7.362	1	.007
[Nonretweet mention of other user=1.00]	0[a]	.	.	.	.	.	.
(Scale)	1						

Dependent Variable: Tweet contains headline or headline excerpt only
Model: (Intercept), Article generalisation in headline, Article headline character count, Nonretweet, Retweet mentioning publication account, Nonretweet mention of other user
a. Set to zero because this parameter is redundant.

Tweet contains full headline (including with other content)

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	-4.558	.6522	-5.836	-3.280	48.839	1	.000
[Article "science" in headline=.00]	1.181	.4348	.329	2.033	7.379	1	.007
[Article "science" in headline=1.00]	0[a]	.	.	.	.	.	.
Article headline character count	.023	.0049	.014	.033	23.321	1	.000
[Article published on Guardian.co.uk=.00]	.622	.2668	.099	1.145	5.435	1	.020
[Article published on Guardian.co.uk=1.00]	0[a]	.	.	.	.	.	.
[Article published on Telegraph.co.uk=.00]	.560	.2720	.027	1.093	4.233	1	.040
[Article published on Telegraph.co.uk=1.00]	0[a]	.	.	.	.	.	.
[Article published on Independent.co.uk=.00]	1.179	.2935	.604	1.755	16.141	1	.000
[Article published on Independent.co.uk=1.00]	0[a]	.	.	.	.	.	.
[Retweet mentioning publication account=.00]	1.337	.2128	.920	1.754	39.442	1	.000
[Retweet mentioning publication account=1.00]	0[a]	.	.	.	.	.	.
[Nonretweet=.00]	.415	.1701	.082	.749	5.958	1	.015
[Nonretweet=1.00]	0[a]	.	.	.	.	.	.
[Nonretweet mention of other user=.00]	-1.334	.3302	-1.981	-.687	16.331	1	.000
[Nonretweet mention of other user=1.00]	0[a]	.	.	.	.	.	.
[Tweet originated from mobile device=.00]	-.274	.1354	-.540	-.009	4.106	1	.043
[Tweet originated from mobile device=1.00]	0[a]	.	.	.	.	.	.
(Scale)	1						

Dependent Variable: Tweet contains full headline (including with other content)
Model: (Intercept), Article "science" in headline, Article headline character count, Article published on Guardian.co.uk, Article published on Telegraph.co.uk, Article published on Independent.co.uk, Retweet mentioning publication account, Nonretweet, Nonretweet mention of other user, Tweet originated from mobile device
a. Set to zero because this parameter is redundant.

Tweet refers to article lead

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	.098	.3128	-.515	.711	.098	1	.755
Article headline character count	.021	.0048	.011	.030	18.608	1	.000
(Scale)	1						

Dependent Variable: Tweet refers to article lead
Model: (Intercept), Article headline character count

Tweet refers to article content outside headline and lead

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	2.597	.4017	1.810	3.384	41.799	1	.000
[Article generalisation in headline=.00]	-1.127	.3935	-1.898	-.355	8.198	1	.004
[Article generalisation in headline=1.00]	0[a]	.	.	.	.	.	.
[Retweet not mentioning publication account=.00]	.696	.2684	.170	1.222	6.729	1	.009
[Retweet not mentioning publication account=1.00]	0[a]	.	.	.	.	.	.
(Scale)	1						

Dependent Variable: Tweet refers to article content outside headline and lead
Model: (Intercept), Article generalisation in headline, Retweet not mentioning publication account
a. Set to zero because this parameter is redundant.

Tweet contains user personal comment

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	-2.934	.6653	-4.238	-1.630	19.445	1	.000
[Article headline highlights study limitations=.00]	1.031	.5226	.006	2.055	3.890	1	.049
[Article headline highlights study limitations=1.00]	0[a]	.	.	.	.	.	.
[Article written by science specialist=.00]	-.318	.1453	-.603	-.033	4.788	1	.029
[Article written by science specialist=1.00]	0[a]	.	.	.	.	.	.
[Retweet not mentioning publication account=.00]	1.182	.1994	.791	1.573	35.135	1	.000

[Retweet not mentioning publication account=1.00]	0[a]	.	.	.	.	.	.
[Nonretweet=.00]	.887	.1539	.585	1.189	33.195	1	.000
[Nonretweet=1.00]	0[a]	.	.	.	.	.	.
[Nonretweet mention of other user=.00]	.779	.2737	.243	1.316	8.106	1	.004
[Nonretweet mention of other user=1.00]	0[a]	.	.	.	.	.	.
[Tweet generated from Twitter reply button=.00]	1.052	.2827	.498	1.606	13.849	1	.000
[Tweet generated from Twitter reply button=1.00]	0[a]	.	.	.	.	.	.
(Scale)	1						

Dependent Variable: Tweet contains user personal comment
Model: (Intercept), Article headline highlights study limitations, Article written by science specialist, Retweet not mentioning publication account, Nonretweet, Nonretweet mention of other user, Tweet generated from Twitter reply button
a. Set to zero because this parameter is redundant.

Tweet indicates user uncritical acceptance of information presented

Parameter Estimates

			95% Wald Confidence Interval		Hypothesis Test		
Parameter	B	Std. Error	Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	1.815	.6938	.456	3.175	6.847	1	.009
[Article headline highlights study limitations=.00]	-1.227	.6413	-2.484	.030	3.662	1	.056
[Article headline highlights study limitations=1.00]	0[a]	.	.	.	.	.	.
[Article headline highlights conflict with other information=.00]	-1.263	.2614	-1.775	-.751	23.343	1	.000
[Article headline highlights conflict with other information=1.00]	0[a]	.	.	.	.	.	.
[Retweet not mentioning publication account=.00]	-.539	.2331	-.996	-.083	5.354	1	.021
[Retweet not mentioning publication account=1.00]	0[a]	.	.	.	.	.	.
[Nonretweet=.00]	-.633	.2670	-1.157	-.110	5.627	1	.018
[Nonretweet=1.00]	0[a]	.	.	.	.	.	.
[Tweet generated from Twitter reply button=.00]	-1.092	.3381	-1.755	-.429	10.429	1	.001
[Tweet generated from Twitter reply button=1.00]	0[a]	.	.	.	.	.	.
(Scale)	1						

Dependent Variable: Tweet indicates user uncritical acceptance of information presented
Model: (Intercept), Article headline highlights study limitations, Article headline highlights conflict with other information, Retweet not mentioning publication account, Nonretweet, Tweet generated from Twitter reply button
a. Set to zero because this parameter is redundant.

Retweet not mentioning publication account

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	4.806	.5995	3.631	5.981	64.274	1	.000
[Tweet user ID appears more than once in sample=0]	-1.120	.2623	-1.635	-.606	18.239	1	.000
[Tweet user ID appears more than once in sample=1]	0[a]	.	.	.	.	.	.
[Tweet originated from mobile device=.00]	1.486	.1339	1.223	1.748	123.054	1	.000
[Tweet originated from mobile device=1.00]	0[a]	.	.	.	.	.	.
Tweet character count	-.029	.0039	-.037	-.022	56.775	1	.000
[Tweet contains full headline (including with other content)=.00]	-.513	.1472	-.802	-.225	12.160	1	.000
[Tweet contains full headline (including with other content)=1.00]	0[a]	.	.	.	.	.	.
(Scale)	1						

Dependent Variable: Retweet not mentioning publication account
Model: (Intercept), Tweet user ID appears more than once in sample, Tweet originated from mobile device, Tweet character count, Tweet contains full headline (including with other content)
a. Set to zero because this parameter is redundant.

Retweet mentioning publication account

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	1.273	.1791	.922	1.624	50.553	1	.000
[Tweet user ID appears more than once in sample=0]	-.677	.1507	-.972	-.381	20.168	1	.000
[Tweet user ID appears more than once in sample=1]	0[a]	.	.	.	.	.	.
[Tweet originated from mobile device=.00]	-.297	.1389	-.569	-.025	4.575	1	.032
[Tweet originated from mobile device=1.00]	0[a]	.	.	.	.	.	.
[Tweet contains full headline (including with other content)=.00]	1.169	.1715	.833	1.505	46.447	1	.000
[Tweet contains full headline (including with other content)=1.00]	0[a]	.	.	.	.	.	.
(Scale)	1						

Dependent Variable: Retweet that mentions publication account
Model: (Intercept), Tweet user ID appears more than once in sample, Tweet originated from mobile device, Tweet contains full headline (including with other content)
a. Set to zero because this parameter is redundant.

Nonretweet

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	.351	.1716	.015	.688	4.189	1	.041
[Tweet user ID appears more than once in sample=0]	1.085	.1406	.809	1.360	59.570	1	.000
[Tweet user ID appears more than once in sample=1]	0[a]	.	.	.	.	.	.
[Tweet originated from mobile device=.00]	-1.170	.1170	-1.399	-.941	99.935	1	.000
[Tweet originated from mobile device=1.00]	0[a]	.	.	.	.	.	.
[Tweet contains full headline (including with other content)=.00]	-.501	.1228	-.742	-.260	16.630	1	.000
[Tweet contains full headline (including with other content)=1.00]	0[a]	.	.	.	.	.	.
(Scale)	1						

Dependent Variable: Not a retweet dummy
Model: (Intercept), Tweet user ID appears more than once in sample, Tweet originated from mobile device, Tweet contains full headline (including with other content)
a. Set to zero because this parameter is redundant.

Tweet generated from Twitter reply button

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	2.749	.1954	2.366	3.132	197.816	1	.000
[Tweet originated from mobile device=.00]	1.573	.2651	1.054	2.093	35.217	1	.000
[Tweet originated from mobile device=1.00]	0[a]	.	.	.	.	.	.
[Tweet contains full headline (including with other content)=.00]	-.686	.2823	-1.239	-.132	5.901	1	.015
[Tweet contains full headline (including with other content)=1.00]	0[a]	.	.	.	.	.	.
(Scale)	1						

Dependent Variable: Tweet generated from Twitter reply button
Model: (Intercept), Tweet originated from mobile device, Tweet contains full headline (including with other content)=

Appendix E: Article characteristics descriptive statistics

Article generalisation

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	111	80.4	80.4	80.4
	Yes, headline only	3	2.2	2.2	82.6
	Yes, headline and lead	20	14.5	14.5	97.1
	Yes, lead only	4	2.9	2.9	100.0
	Total	138	100.0	100.0	

Article advice

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	70	50.7	50.7	50.7
	Yes, headline only	7	5.1	5.1	55.8
	Yes, headline and lead	54	39.1	39.1	94.9
	Yes, lead only	7	5.1	5.1	100.0
	Total	138	100.0	100.0	

Article policy recommendation

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	131	94.9	94.9	94.9
	Yes, headline and lead	4	2.9	2.9	97.8
	Yes, lead only	3	2.2	2.2	100.0
	Total	138	100.0	100.0	

Article highlights conflict with other information

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	124	89.9	89.9	89.9
	Yes, headline only	3	2.2	2.2	92.0
	Yes, headline and lead	3	2.2	2.2	94.2
	Yes, lead only	8	5.8	5.8	100.0
	Total	138	100.0	100.0	

Article highlights study limitations

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	134	97.1	97.1	97.1
	Yes, headline and lead	3	2.2	2.2	99.3
	Yes, lead only	1	.7	.7	100.0
	Total	138	100.0	100.0	

Article mention of "science"

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	122	88.4	88.4	88.4
	Yes, headline only	4	2.9	2.9	91.3
	Yes, lead only	12	8.7	8.7	100.0
	Total	138	100.0	100.0	