

# **The Role of Online Discussion Forums During a Public Health Emergency**

**Jennifer Ann Cole**

**Thesis submitted for the degree of  
Doctor of Philosophy**



**Royal Holloway  
University of London**

**2017**



# Declaration of Authorship

I, Jennifer Ann Cole, hereby declare that this thesis and the work presented in it is entirely my own. Where I have consulted the work of others, this is always clearly stated.

Signed:

Dated:



# Abstract

During a public health emergency, access to professional healthcare may be constrained. Online discussion forums provide an alternative – access to a ‘Wise Crowd’ from which Collective Intelligence may emerge – but lie outside of traditional quality control structures.

This thesis examines whether certain characteristics of such platforms encourage and signpost higher quality information, and what utility this may offer during a Public Health Emergency of International Concern.

It synthesises results from three separate studies: interviews with Ebola witnesses during the 2014-16 West Africa outbreak; doctors’ assessments of the quality of information provided in response to questions asked on health discussion forums, including r/ebola on Reddit.com, dedicated to discussing Ebola; and an investigation of online health community moderators by non-participant observation, interviews, and access to private areas of discussion sites not accessible to general users.

The research goes beyond previous work in three major ways:

First, analysis of health-seeking behaviour presented in this thesis identifies three separate stages of risk perception and individual concern during an outbreak with different types of questions asked in each stage, each requiring different platform characteristics.

Second, assessment of the quality of the health advice in major online forums is determined in this study according to the judgements of doctors; this differs

from previous research that used indirect assessment. Results suggests that such forums can provide signposts to good quality information.

Third, the investigation of the role of teams of volunteer moderators of reddit discussion forums identifies and analyses the task-components and skill-mix required to set up and manage a discussion forum during a serious disease outbreak. This points to a larger role for technical skills and forum experience than has previously been identified.

The conclusions of this thesis suggest ways in which online platforms may facilitate better health information exchange during future serious disease outbreaks.

# Acknowledgements

My PhD has been made possible by several individuals and organisations, all of whom deserve full acknowledgement.

I have been funded in full by Royal Holloway University of London through a Reid Scholarship in Health, the Human Body and Behaviour, an innovative attempt to promote interdisciplinary research, which in this case involved the School of Biological Sciences, the Department of Computer Science, and the Department of Geography. The ability to draw on the expertise, interests and resources of all three departments has made the study stronger and more far-reaching than could have been achieved from a single disciplinary approach.

The most important contribution made by the departments were of course my supervisors – Professor Chris Watkins from the Department of Computer Science along with Professor Dorothea Kleine and Dr Alasdair Pinkerton from the Department of Geography. Each has guided, critiqued and shaped my approach throughout and has helped me to hone and focus the results.

I would also like to acknowledge the reddit moderators who not only gave their time to answer my questions, but also provided access to private areas of their forums beyond those accessible to general users, including statistics, data and report logs that proved invaluable to my research. So much of what they do is invisible to the reddit users who benefit from their efforts: they work incredibly hard for little reward.

I would also like to acknowledge those people who have had a less direct involvement with my study, but without whom it would not have been possible. The first of these is Elizabeth Jesser Reid (1789-1866). Her legacy

provided my scholarship. More importantly, she founded Bedford College, the first higher education institution for women in Great Britain and opened the door to higher education for every woman who came after her.

Secondly, I would like to acknowledge Aaron Swartz (1986-2013), the American computer programmer and activist who was an early investor in reddit and who guided its development towards the provision of news and information. A strong campaigner for open access, he took his own life while awaiting trial for allegedly downloading articles from the digital academic library JSTOR with the intention of sharing them on illegal file-sharing websites. Reddit is an incredible free and open information resource that would not be what it is without his vision, which has shaped it far more than his financial investment.

Finally, I would like to acknowledge my late in-laws, Michael and Bridget Hand, who died in March 2014 shortly after my study began and whose financial legacy enabled me to switch from part-time to full-time study and, in all probability, to complete the course of study at all. Higher Education is the realm of the financially comfortable, and I would ask anyone who reads this thesis to continue to do whatever they can to ensure access is open to all. The idea that it is at present is a fallacy that should never go unchallenged.

## CONTENTS

---

### **I: INTRODUCTION**

- 1.1 Introduction
- 1.2 Communication during public health emergencies
- 1.3 The research question
- 1.4 Justification for the research
- 1.5 Health information seeking online
- 1.6 Public Health Emergencies of International Concern
- 1.7 Ebola Virus Disease: 2014-16 West Africa outbreak
- 1.8 r/ebola: a case study
- 1.9 Gaps in the current understanding
- 1.10 Contribution to knowledge
- 1.11 Scope and structure of the study
- 1.12 Limitations of the study
- 1.13 Conclusions

### **2: THEORETICAL FRAMEWORK AND LITERATURE REVIEW**

- 2.1 Introduction
- 2.2 Technology affordances
- 2.3 Collective Intelligence
- 2.4 How good is good enough?
- 2.5 The computer as disruptive technology
- 2.6 Pierre Lévy's Theory of Collective Intelligence

2.7 The Collective Intelligence Genome

2.8 Conclusions

### **3: METHODOLOGY**

3.1 Introduction

3.2 Phase I: Interviews with Ebola witnesses

3.3 Phase II: Quality of information in online health forums

3.4 Phase III: Observation of reddit and interviews with moderators

3.5 Summary of data collected

### **4: REDDIT – ‘THE FRONT PAGE OF THE INTERNET’**

4.1 Introduction

4.2 Structure of reddit

4.3 Voting on reddit

4.4 Ordering of information on reddit

4.5 Discussion

4.6 Conclusions

### **5: INFORMATION REQUIREMENTS DURING A PHEIC**

5.1 Information requirements during a PHEIC

5.2 Questions during the 2014-16 *Zaire ebolavirus* outbreak

5.3 Context Shift: Far at Risk, Near at Risk, Real at Risk

5.4 Context Shift on r/ebola

5.5 Community Shift on r/ebola

5.6 Post-risk: Context and Community Shift

5.7 Discussion

5.8 Conclusions

## **6: INFORMATION RETURN: ANSWERS FROM REDDIT**

6.1 Introduction

6.2 The relationship between trust and quality

6.3 Credibility on r/ebola

6.4 Searching for answers – the typical information-seeking journey

6.5 Quality of information on reddit health forums

6.6 Phase II-A: Investigation of information quality in online health forums (pilot study)

6.7 Phase II-B: Investigation of information quality in r/ebola

6.8 Phase III: Observation of reddit and interviews with moderators

6.9 Conclusions

## **7: WHO ARE THE MODERATORS AND WHAT DO THEY DO?**

7.1 Introduction

7.2 Definition of a reddit moderator used in the study

7.3 Moderation roles

7.4 Technical skills: using the moderation tools available

7.5 Becoming a moderator

7.6 Moderation teams

7.7 Moderation requirements and Context/Community Shift

7.8 Moderator metacommunities – reachback into the reddit crowd

7.9 Conclusions

## **8: HOW SUBREDDITS SCALE UP IN A PHEIC**

8.1 Introduction

8.2 The PHEIC subreddits

8.3 Growth of r/ebola

8.4 Genesis of r/ebola

8.5 From r/ebola to r/zika

8.6 When do surges occur?

8.7 Discussion and conclusions

## **9: DISCUSSION AND CONCLUSIONS**

9.1 Lessons identified from r/ebola

9.2 Limitations of the study

9.3 The information journey and dimensions of risk

9.4 Context Shift and Community Shift

9.5 Credibility and trust on Reddit

9.6 The Expert Moderator

9.7 Conclusions

## **10: REFERENCES**

## FIGURES

---

- FIG 1:** *Proposed model for how users judge the credibility of online information*
- FIG 2:** *Technology Acceptance Model*
- FIG 3:** *The Choice Framework*
- FIG 4:** *The Collective Intelligence Genome*
- FIG 5:** *A screenshot of the message posted on Liberia Expats Googlegroup*
- FIG 6:** *An example of the message posted on reddit health forums*
- FIG 7:** *A typical example of a reddit Front Page*
- FIG 8:** *The Front Page contains several different elements*
- FIG 9:** *A series of link posts from the subreddit r/zika\_en.*
- FIG 10:** *A self-post asking for advice regarding the Zika virus*
- FIG 11:** *Reddit registered users can comment on posts*
- FIG 12:** *Votes are cast by clicking on the up or down arrow*
- FIG 13:** *Clicking on the username of poster brings up their profile*
- FIG 14:** *Users can click through to subreddits using the menu bar*
- FIG 15:** *Search results page showing subreddits and posts returned*
- FIG 16:** *The order in which posts are displayed on reddit is determined by the hotscore algorithm*
- FIG 17:** *How (r) affects the position of posts on reddit*
- FIG 18:** *Comments below a threshold are hidden*

- FIG 19:** *Tabs at the top of the page*
- FIG 20:** *Comments are default ordered by 'best'*
- FIG 21:** *Timeline of Ebola cases*
- FIG 22:** *Source of all posts on r/ebola, July 2014*
- FIG 23:** *Link posts on r/ebola from early in the outbreak, in July 2014*
- FIG 24:** *Each unique visitor viewed r/ebola twice on average in June 2014*
- FIG 25:** *Total number of link-posts and self-posts on r/ebola*
- FIG 26:** *Icons distinguish between link posts and self-posts*
- FIG 27:** *Timeline for the r/ebola subreddit showing subscriber growth*
- FIG 28:** *There are two distinct spikes in subscriber growth [on r/ebola]*
- FIG 29:** *A search for Ebola-related subreddits*
- FIG 30:** *Significant discussion on Ebola has also taken place outside of Ebola subreddits*
- FIG 31:** *Posts on r/ebola display several signposts*
- FIG 32:** *Results returned for a search inside r/ebola on 'airborne'*
- FIG 33:** *The only story about airborne transmission of Ebola from a source flaired as science/medicine takes a clear position this is not likely to happen*
- FIG 34:** *Only three posts about Ebola becoming airborne are found on r/science*
- FIG 35:** *The same search, on 'ebola airborne' on r/worldnews*

- FIG 36:** *A higher percentage of respondents rated information as highly sensible, appropriate and useful than highly complete or accurate*
- FIG 37:** *Doctors considered a significant proportion of the information to be mostly accurate and mostly complete*
- FIG 38:** *The non-doctor respondents were more likely to award mid-range scores*
- FIG 39:** *Correlation between each doctor's score with the mean of other doctors*
- FIG 40:** *A post implying that the Ebola virus could go airborne*
- FIG 41:** *Some posts submitted to the r/zika subreddit subscribe to conspiracy theories*
- FIG 42:** *Moderators of a subreddit are listed on its front page*
- FIG 43:** *Proportion of moderators' time spent on different tasks*
- FIG 44:** *The moderation tools box*
- FIG 45:** *An excerpt from the automoderator configuration for r/zika*
- FIG 46:** *Another excerpt from the automoderator configuration*
- FIG 47:** *There is a general trend for an increase in the number of moderators as the number of subscribers increases*
- FIG 48:** *Distribution of tasks recorded on the moderation log of r/zika*
- FIG 49:** *Growth of subscribers on r/cystic fibrosis*
- FIG 50:** *Growth of subscribers on r/diabetes*
- FIG 51:** *Growth of subscribers on r/asthma*

- FIG 52:** *r/worldnews has not altered its design from the default setting*
- FIG 53:** *r/books design has been modified to be more graphic*
- FIG 54:** *Subscriber growth on r/ebola*
- FIG 55:** *Subscriber growth on r/zika*
- FIG 56:** *Page views by day for r/ebola from September-October 2014.*
- FIG 57:** *Timeline for r/ebola's subscriber growth*
- FIG 58:** *Spikes in activity on the r/zika subreddit have been less pronounced*
- FIG 59:** *An entry from r/ebola modmail logs (above) showing discussion on how to assign filters to posts.*
- FIG 60:** *The system on r/technology*
- FIG 61:** *List of the flairs available on r/ebola as of 27<sup>th</sup> July 2016*
- FIG 62:** *A discussion started on r/ebola on 2<sup>nd</sup> October 2014*
- FIG 63:** *The moderators' reply to a discussion on the removal of posts*
- FIG 64:** *An advert that appeared on r/ebola at the height of the outbreak.*
- FIG 65:** *Message from the moderator of r/zika to the r/ebola moderation team*
- FIG 66:** *Pattern of new subscriber growth on r/ebola*
- FIG 67:** *Subscriber growth on r/starwars*

## TABLES

---

- TABLE 1:** *The main research question and sub-research questions addressed in this thesis*
- TABLE 2:** *The Collective Intelligent Genome and its Variants (after Malone, Laubacher and Dellacros, 2009)*
- TABLE 3:** *Demographics of participants in the Phase II-B study*
- TABLE 4:** *Forums included in this study and the number of moderators on each*
- TABLE 5:** *Qualitative and quantitative data collected and analysed*
- TABLE 6:** *Elements of the reddit Front Page*
- TABLE 7:** *Elements of a post*
- TABLE 8:** *Categorisation of subreddit types*
- TABLE 9:** *How the number of votes cast affects the 'best' ordering.*
- TABLE 10:** *The best score places some comments ...*
- TABLE 11:** *Interview dates and level of risk*
- TABLE 12:** *Adapted from Table 3, Far at Risk, Real at Risk, Near at Risk, in Cole and Watkins (2015).*
- TABLE 13:** *Source of link posts on r/ebola in July 2014*
- TABLE 14:** *Interviewees wanted information explained more clearly*
- TABLE 15:** *Preferred sources of information at different stages of the outbreak*

- TABLE 16:** *Information sources and perceptions about them (adapted from Cole and Watkins, 2015)*
- TABLE 17:** *The number of unique visitors, page views and views per unique visitor to r/ebola increased between June and October 2014*
- TABLE 18:** *Link posts and self-posts on r/ebola*
- TABLE 19:** *Reddit moderators of r/ebola interviewed during this study*
- TABLE 20:** *Difference in questions asked by far at-risk, near at-risk and actual at-risk groups*
- TABLE 21:** *The overall score of each thread as given by each assessor*
- TABLE 22:** *Correlations of each doctor's scores*
- TABLE 23:** *Table showing voting behaviour on the r/ebola subreddit*
- TABLE 24:** *Treatment by the r/ebola community of answers to which none of the doctors surveyed gave a 'good' rating.*
- TABLE 25:** *How the different hierarchical levels of reddit are able to influence the quality of information on the site*
- TABLE 26:** *Skills needed by reddit moderators*
- TABLE 27:** *Moderation tools and functions available*
- TABLE 28:** *How those interviewed for this study became moderators*

## APPENDICES

---

- Appendix I**      *Questions for People Stationed in Ebola-Affected Regions of the World* (1 page) from PHASE I: Interviews with Ebola Witnesses. Interview guide.
- Appendix II**      *Quality of Online Health Information: Discussion Threads for Assessment* (7 pages) from PHASE II-A: Investigation of information quality in online health forums (pilot study). Instructions for participants.
- Appendix III**      *Discussion Threads for Assessment: Results* (4 pages), from PHASE II-A: Investigation of information quality in online health forums (pilot study). Results data.
- Appendix IV**      *Health Information Seeking During the 2014-16 Ebola Outbreak: A Study on the Quality of Health Information* (17 pages) from PHASE II-B: Investigation of information quality in r/ebola. Instructions for survey participants.
- Appendix V**      *Investigation of information quality in r/ebola: Results data* (2 pages) from PHASE II-B Investigation of information quality in r/ebola. Results data.
- Appendix VI**      *Questions for moderators of [the reddit xxx forum]* (4 pages) from PHASE III: Observation of reddit and interviews with moderators. Interview Guide.



## **I: INTRODUCTION**

---

### **1.1 Introduction**

Imagine the sudden emergence of a new disease, previously unknown to medical science, which infects more than half the UK population. With no natural immunity present, it kills around 750,000 people, or just under 1% of the population. This equates to 2-3 children in every average-sized UK primary school, 5-15 children in every average-sized UK secondary school (Department of Education, 2015). In every street in the UK, in every tower block, households would be affected, there would be at least some deaths. By the end of the outbreak, everyone would know someone personally – a close friend, relative or co-worker – who had succumbed to the disease. These are the planning assumptions on which the UK Government prepares for a serious outbreak of pandemic flu, or other potential public health emergency (Cabinet Office, 2013). But what specific challenges does a public health emergency raise, and are we better placed in the 21<sup>st</sup> century to address one than we were in the 20<sup>th</sup>, when the 1918-19 Spanish Influenza pandemic caused upwards of 20 million deaths worldwide (Glezen, 1996)?

The idea for this thesis emerged from an interdisciplinary conference on the resilience challenges of serious infectious disease that took place in February 2013 and brought together academics and policy makers to discuss not only which infectious diseases to plan for, but also what measures might be considered during such events to slow or stop the spread of the outbreak. Funded by the Science and Technology Facilities Council (STFC) and convened by the Royal United Services Institute for Defence and Security Studies (RUSI), the event deliberately brought together epidemiologists, microbiologists, risk analysts, behavioural psychologists, mathematical modellers, medics and health planners to discuss how different disciplines

would approach the challenges. As part of the proceedings, in my role as Senior Research Fellow, Resilience and Emergency Management at RUSI and as a member of the UK Cabinet Office National Warning and Informing Group Steering Committee (an independent body that advises government on communication with the public in advance of and during emergencies), I ran a seminar discussion on how policymakers might better understand public concerns and, through this understanding, influence public behaviour during a future pandemic. My position as a policy researcher, with a professional background in journalism, an academic background in biological anthropology and experience of working through the 2009 influenza pandemic – colloquially known as ‘Swine Flu’ – gave me a strong interest in how public health messages might be used to influence behaviour and reduce the spread and impact of the disease. The deliberately broad scope of the conference highlighted the complex and challenging threat posed by a potential outbreak of a serious, novel – and at first likely to be difficult to treat or incurable – disease. The output from the event was a series of research ideas that could be further explored by the participants. I decided to take one of those identified research needs – understanding how promoting public discussion using digital communication technologies might influence the outbreak (Cole and Watkins 2013, p111) – and make it the basis for this thesis.

In this chapter, I will explain how I approached the literature, how this influenced my research design and methods and how the horrific 2014-16 Ebola outbreak in West Africa offered me the unique opportunity to study an actual unfolding outbreak of serious infectious disease in real time.

## **1.2 Communication during public health emergencies**

Emerging infectious diseases sit right at the heart of the UK Cabinet Office’s National Risk Register (Cabinet Office, 2015), the official UK Government

assessment of significant potential risks to the UK public. The risk from emerging infectious diseases is classified as being at mid-range for both likelihood (between a 1 in 20 and 1 in 200 predicted chance of happening in the next five years) and likely impact (rated 3 on a scale of 1-5), while pandemic influenza is singled out and given the top rating, 1, for likely impact. Significant UK government planning has gone into deciding how the risk from such disease outbreaks should be categorised, and what measures should be taken to prepare for them, set out in documents such as Public Health England's *Communicable Disease Outbreak Management: Operational Guidance* (PHE 2014), and the Foresight Report *Infectious Diseases: Preparing for the Future* (Brownlie et al, 2006).

While these documents focus on recording, identifying and treating those infected in a potential outbreak, they do not seem to consider how the public might be able to take a proactive part in preventing or slowing its spread. This led me to wonder if it might be possible to use communication channels, including social media and peer-to-peer communication, to encourage the uptake of protective behaviours such as social distancing, or to enable those in affected areas to share information on, for example, specific streets or transport routes where cases were concentrated.

UK government guidance on communicating with the public aligns with the World Health Organization's *Outbreak Communication Planning Guide* (WHO, 2008). This states (p4):

*“[P]ro-active communication encourages the public to adopt protective behaviours, facilitates heightened disease surveillance, reduces confusion and allows for better allocation of resources – all of which are necessary for an effective response.”*

A pro-active communication strategy is seen to be particularly important in the early stages of the outbreak, when the control opportunities are highest and it may be possible to slow to spread or to contain small, localised outbreaks before the disease becomes endemic.

Mathematical modelling of theoretical outbreaks (Funk et al, 2009 and 2010; Funk and Jansen, 2013) also suggests that in the early stages, human behaviour may have a significant role to play in limiting outbreak spread: effective, self-initiated messages shared through peer-to-peer networks in the proximity of new cases may help to contain those outbreaks at a local level. Yet there is little evidence that considerations of peer-to-peer information sharing have been made in the UK Government planning assumptions. The PHE document provides guidance on how communications should be managed, but the focus on communication is one way: out from the authorities to the passively receiving public. Nguyen-Van-Tam and Sellwood (2013a) list 16 non-pharmaceutical/vaccine interventions for limiting the spread of pandemic flu in three categories (international travel measures; personal protection and social distancing) but do not explore how peer-to-peer platforms might support promotion and implementation of these interventions, for instance by quickly alerting members of a social group to the first case recorded in their network and encouraging others to immediately self-distance.

I found little evidence in the literature on pandemic preparedness to suggest that peer-to-peer channels are considered within existing plans. PHE's media strategy outlines the type of information that might be communicated to the public through professional media – such as the number and location of cases – and considers communicating updates on such information via social media networks such as Twitter and Facebook. This considers only how information will be pushed out from the public health authority, however, not how the public might communicate their concerns, questions or fears back to the

authorities, their GP or local hospital, or to how they might discuss issues between themselves so that they might self-organise social distancing strategies; share advice and experience on how to cope with the outbreak; or ask questions that will enhance their understanding of what is happening. The Department of Health's *UK Pandemic Influenza Communications Strategy 2012* (Department of Health, 2012), also focusses only on the push of information from official agencies to the public. The Foresight Report (Brownlie et al, 2006) considers how the authorities might best communicate to the public the need for increased disease surveillance which may (or may be seen to) invade privacy, and would thus need to be supported by justifications and additional information, but does not consider any mechanism through which the public might ask for clarification or be able to discuss proposals.

U.S. government planning also recognises effective public communication as a crucial component of the response to health emergencies (Reynolds and Seeger, 2005; Reynolds and Quinn, 2008) without considering how interactive communication might play a part. The U.S. Center for Disease Controls' *Pre-Event Message Development Project* (PEMDP) (Wray et al, 2008), a comprehensive study on information needs before, during and after health emergencies, makes no mention of how the public might seek to communicate with one another through peer-to-peer networks.

My own experience as a policy researcher, particularly from the Swine Flu epidemic of 2009-10, suggested that a better understanding of the role peer-to-peer communication could play may benefit health emergency planning. During the pandemic, I worked with *The Guardian* to help its journalists understand the science and epidemiology of the disease, and to shape its coverage to best answer concerns communicated by the newspaper's readers. The questions asked tended to be different to those being answered by the Department of Health and the NHS, such as when might the outbreak be over,

so that workers would no longer have to cover staff shortages caused by illness, and could plan to take holiday again? Why did cases not seem to be as severe as had been predicted? Why was this type of flu different to others?

These were not the type of questions one would typically book a consultation with a GP to ask, nor that the NHS flu helpline – set up to help diagnose people who thought they were displaying symptoms – was best placed to answer, but they did highlight concerns. This led me to consider how such questions could be answered, and what type of platforms might facilitate this.

### **1.3 The research question**

The considerations outlined above shaped my main research question:

*In what way(s) do the characteristics of online discussion forums facilitate or hinder health information seeking online?*

I began to think about how a platform might be best configured to ensure that it works well and that people will want to use it. This led me to first explore literature on Technology Affordances, which explains what technology enables us to do (Gibson, 1979; Norman, 1986, 1988, 1999), and from human-computer interaction (HCI), which explains how we interact with that technology, including user-centered design (Davis, 1989) and the Technology Acceptance Model (Davis, 1986; Davis et al, 1989). I used these literatures – described in more detail in Chapter 2 – to understand what characteristics a platform might need to encourage people to use it.

Secondly, I considered how people involved in such an event might discuss their situation and attempt to formulate answers collectively. I next looked at literature on Collective Intelligence – the role of computer networks in linking

people together to communicate in ways that are not possible without those networks (Hiltz and Turoff, 1978; Lévy, 1995, 1997; Malone, 2006 and 2008). I compared this with literature on human cooperation and group working offline (Galton, 1907; Surowiecki, 2005; Tomasello, 2009a, 2014) from anthropology and organisational theory, before coming back to consider how the online environment might enhance collective endeavour (Lévy, 1995, 1997; Woolley et al, 2015; Broadbent and Galloti, 2015). These literatures helped me to understand how the users of a system interact with it and with one another across it. I explore these literatures, and my approach to them, further in Chapter 2, while in Chapter 3, I will also explain how they led me to consider whether online peer-to-peer networks display appropriate technology affordances to meet the information seekers' needs and how this influenced the methods I used to explore my research questions.

#### **1.4 Justification for the research**

My initial exploration of the literature suggested that while public health messaging is acknowledged as important during public health emergencies, little consideration has been given to any form of communication other than a push out from public health authorities to the public, who are expected to receive the information provided, and perhaps act on it, but not to enter a two-way discussion. Unless a member of the public is infected and requires diagnosis or treatment, there is no provision for them to discuss concerns or to ask for advice that might help them to deal with the situation in which they find themselves. Studies of health-seeking behaviour consistently point to a qualified medical practitioner as being health information seekers' first preference (Hesse et al, 2005; Wray et al 2008; Cole and Watkins, 2015; Diviani et al, 2015), but if such access is not readily available during a public health emergency, those with questions may look for alternatives. A lack of access to health information and advice at the usual or preferred point of entry

(one's GP), may increase pressure on NHS resources further up the chain – 9.8% of patients who report that it is not easy to get through to their GP's surgery on the telephone go to an A&E walk-in centre instead (NHS, 2013). Alternative provision may become extremely important during a public health emergency that is stretching resources, especially if it could help deal with the 'worried well' – those who have not been exposed or infected but worry they might be (Hyer and Covello, 2005) – freeing up healthcare providers to deal with their normal business and those patients who need their help the most.

This influenced my approach by encouraging me to wonder exactly what questions such people might have, and who might be second-best placed, after a doctor, to answer them?

One place people might turn to discuss health concerns is health discussion forum websites. When health information seekers go online, they generally search on a specific condition (Akerkar and Bichile, 2004) and start with a question they want answered (Cole and Watkins, 2015; Diviani et al, 2015). Discussion forums are not only able to answer those questions but also enable discussion around them. This has several potential benefits in health emergencies: the obvious one is as a substitute for the discussions a health information seeker ideally wants to have with their GP.

Such forums might also be able to encourage social distancing, help with surveillance, promote public health messages, and encourage good practice to be shared. They might allow new and potential policies on how to contain the outbreak to be debated (Cole and Watkins, 2013). This may support individuals to shift at least some health-seeking behaviour from a face-to-face consultation with a medically qualified practitioner to the online community, relieving pressure on NHS helplines, GPs surgeries and A&E departments.

This led me to think about how such a platform would need to be configured to help people answer the questions they may have, and what I would need to understand to determine this. To plan an appropriate research programme, I first needed to know the extent to which people currently use online health information platforms and what they use them for.

### **1.5 Health information seeking online**

During the same period as pressure on GPs, hospitals and waiting times has been increasing in the UK, online health-seeking information has also increased dramatically. In 2008, only 18% of UK adults said they looked for health information online, but by 2013 this had risen to 43%, with an increase of 59% amongst the 25-29 age group (ONS, 2015). Health information seeking represented one of the fastest growing areas of Internet use measured by the UK Government during the period 2008-2013. The UK is the second highest country globally for Internet health searches, and one recent survey (Pushdoctor, 2015) rates 'Google my symptoms' as a more common first action than 'Book a Doctor's appointment' or 'Visit a pharmacy for advice'. The same survey estimated that in 2014, the number of health searches carried out in the UK increased by 19%.

This increase should not be surprising considering the huge amount of health information that can be found easily on the Internet. On 10<sup>th</sup> March 2017, a Google search of common health terms returned 257 million hits for 'Diabetes', 343 million hits for 'Pregnancy', 160 million hits for 'HIV', 112 million hits for 'flu' and four million hits for 'Chickenpox'. Webpages found through such searches range from official information put out by Governments and international agencies such as the World Health Organization, to personal blogs by individuals with no apparent medical training or qualifications, however, and while Esquivel et al (2006) have suggested that ideally, health

consumers should be able to access online health information without the need for professional guidance, this requires online resources to be accurate, and for health information consumers to be able to confidently find examples they know they can trust. This will be particularly important if such platforms are to be supported and endorsed by healthcare providers and policymakers. This suggested to me that understanding how users can identify good information is a key challenge, for which we need to consider technology affordances including website characteristics, systems architecture and signposting as well as what helps people to trust information they find online.

The literature on health information seeking I explored suggests that at present, online health information is not well trusted. Many people in the UK currently search for health information online, but not as many as use the Internet to search for other information (Ofcom, 2016): only 20% (an increase from 16% in 2014) of UK adults use the Internet for health each week, while 89% use it for communication. In 2014 (Ofcom, 2014), 37% of UK adults said they search online to ‘find information about health-related issues’ at least quarterly, but this was only half as many as shopped online (66%), and fewer than used it to access public services in general (40%). These figures cover the entire UK population, however, not only those who had a health issue during the period in which the statistics were collected and who therefore had a reason to search, which may account for the relatively low percentages.

Even within online health information seeking, however, peer-to-peer health platforms seem to be particularly underused: combined, online communities, social networking sites, blogs and forums account for only 15% of the online health information landscape (Groselj, 2014). Whitelaw et al (2014) remarked on how few social networking sites were returned using their chosen search criteria during a study of information on birth options available to women in the UK, as such sites are known to be widely used for support by pregnant

women (Dahlen and Homer, 2013). As neither study determined why such forums are underused, rather than simply noting that they are, there is a clear gap in our understanding that needs to be addressed. Is there a case for trying to improve this and, if so, how might that improvement be achieved?

My study does not propose that Internet health discussions forums should become a replacement for face-to-face interaction with qualified healthcare professionals, nor that any source other than qualified healthcare professionals should be the first choice when seeking health information, diagnosis or advice. Instead, it recognizes that during a severe public health emergency, resources may be stretched and the ability to interact face-to-face with a qualified medical professional may be limited or entirely unavailable. I explore whether, under such circumstances, online health discussion forums add *some* value, even if this is not *full* value. Valuable contributions might include online interaction with a qualified healthcare professional who is geographically distanced but who can be communicated with remotely rather than in person, or online interaction with non-medically qualified but experienced individuals, such as others who have experienced the same health condition and who may be able to offer some advice.

If local healthcare resources were to become even more stretched, even interaction with non-medically qualified individuals with little or no expertise or experience but who are nonetheless willing to offer what advice they can, may still be better than nothing. This encouraged me to turn to literature on how decisions are made in challenging circumstances, such as when the subject under consideration is beyond anyone's personal experience (Dalkey, 1969; Linstone and Turoff, 1975 and 2011; Gupta and Clarke, 1996; Ghamari-Tabrizi, 2000) or when resources are constrained (Chambers 1981, 1992; Sen, 1997). From this, I deduced that such decision making has two components: [1] how best to 'pool' partial amounts of information held within a group so

that the group can construct something more complete than any one member would be able to alone, and [2] how to identify which, out of a series of answers amongst which none may be perfect, is better than the others? This led me to also consider literature that explores how group collaboration can shape information, including the Wisdom of Crowds (Galton, 1907; Surowieki, 2005), which identifies that the average guess of a group is more likely to be accurate than that of an individual; The Delphi Method, (Dalkey, 1969), a technique for making group decisions for challenges that no single expert has sufficient knowledge or experience to answer alone; and Participatory Rural Action (Chambers, 1981), in which group interviews help to check and correct information provided by individuals. I use these to help me to understand how health discussion forums can promote accuracy through group discussion and collaboration. I will return to this literature in Chapter 2.

To answer my main research question, I also needed to understand what characteristics online health information platforms have, and how these help or hinder health information seeking. Characteristics that help the process can then be prioritised, while those that hinder it can be mitigated or designed out.

One factor that looms large in the existing literature on online health information seeking is quality. Poor quality information could lead to an incorrect diagnosis, leading the seeker to miss out on appropriate treatment or to self-administer the wrong treatment. The risks associated with inaccurate or misleading health information are therefore higher than for other types of online information (Luo and Najdawi, 2004): there may not be a 'second chance' to choose the right medicine in the same way that a consumer can buy a second watch if their first purchase does not work. A strong criticism levied against 'the Internet' in studies of online health information is that it is an ungoverned space whose information lacks the quality control provided by the editors and peer reviewers of medical textbooks and journals (Akerkar and

Bichile, 2004; Venkatesan et al, 2016) but such studies tend to treat ‘the Internet’ as an homogenous whole, with few considering how the characteristics of some websites might be more conducive to higher quality than others. Most online forums are ‘moderated’, for example, overseen by an individual or group of individuals whose role(s) go at least some way to fulfilling that of an editor or peer reviewer in assuring information quality. Even in unmoderated forums, there is evidence that the community self-polices and self-corrects information posted there (Mursch and Behnke-Mursch, 2003; Esquivel et al, 2006).

The literature also suggests, however, that remarkably little of the information that appears online is of poor quality or likely to be harmful (Eysenbach and Khöler, 2002; Crocco et al 2002; Nölke et al, 2015), though the studies that have explored this have not compared one type of website against another, or one website against another with different characteristics. Furthermore, the methods used to determine ‘quality’ have more often been based on whether the site displays some type of accreditation seal (Burkell, J., 2004; Luo and Najdawi, 2004; Lawrentschuk et al, 2012) or how closely it conforms to medical textbooks (Impicciatore et al, 1997) or information leaflets (Whitelaw et al, 2014) rather than asking doctors to assess if the information is good quality. This informed the design of a pilot study (presented in detail in Chapter 6) in which I asked doctors to rate the quality of information on three separate discussion sites ([www.reddit.com](http://www.reddit.com); [www.patient.co.uk](http://www.patient.co.uk); and [www.musmnet.com](http://www.musmnet.com)), and began to explore what site characteristics might be directly influencing the quality of the information found there.

As well as quality *per se* being an important characteristic, another is whether users trust (or don’t trust) information. Lack of trust is often suggested as a reason for the relatively low incidence of online health information seeking, with individuals lacking confidence in their ability to recognise what is good

information and what is less accurate (Terrace et al, 2001; Akerkar and Bichile, 2004; Diviani et al, 2015). Even highly educated people may have trouble identifying misleading health information: Shon et al (2000) found that participants in their study who had a U.S. college degree were just as likely to believe false information as those without, and van der Vaart et al (2011) could find only slight correlations between education and eHeals, a score of health literacy, which could not be proven to be statistically significant. If a lack of confidence in one's ability to distinguish between accurate and inaccurate information is a reason why online health information platforms are underused, website characteristics that enhance users' ability to quickly locate health information they feel they can trust may be one way to drive up use.

A recent study by Diviani et al (2015) suggests that while people who consider themselves to have lower health literacy are less likely to search online than those who are more confident, when they do they place value on institutional authorship and presence of the author's credentials. This suggests that they self-compensate for an acknowledged lack of expertise by looking for trust markers that give an indication of the information's likely quality. This informed my research design by suggesting that it would be valuable to find out what trust markers might be present in online forums and how they are recognised. This encouraged me to incorporate interviews with health forum users into my design approach and to ask for their opinions, as well as just to observe the exchange of information in the forums. The information they provided is presented and analysed in Chapters 6 and 7.

Understanding what drives trust may be particularly important for how a new system can be configured – or an existing one modified – to provide a suitable platform during a public health emergency. People with little experience of the unfolding health event will need to be confident that any information they find is of high quality. A useful framework is provided by 'The Proposed Model

for 'How Users Judge the Credibility of On-line Information', developed by Wathen & Burkell (2002), shown below:

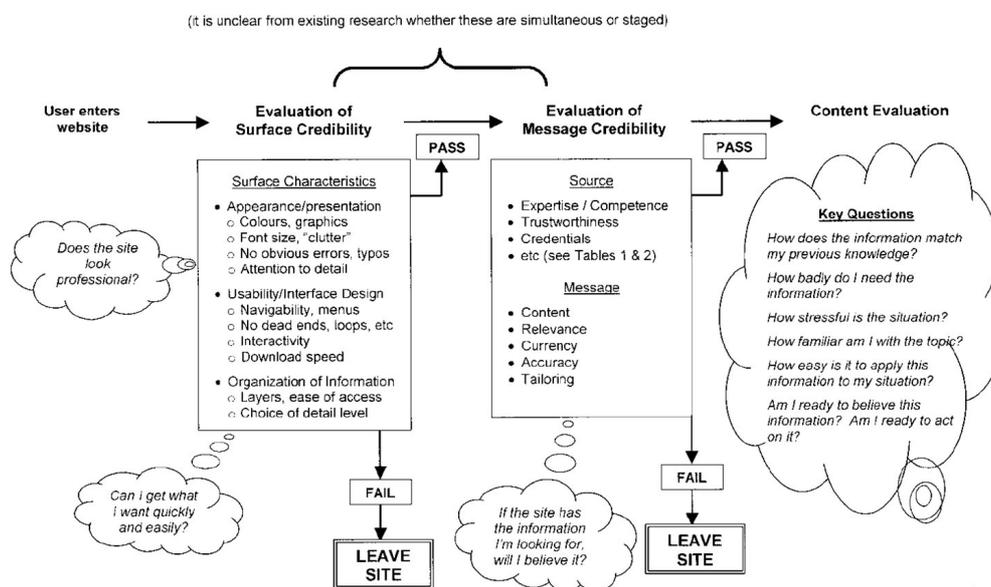


FIG 1: Proposed model for how users judge the credibility of online information (Wathen and Burkell, 2002)

The model considers how factors including source characteristics (who originated the information), channel characteristics (including professional design, good spelling and grammar), medium characteristics (is the website always available? it is easy to navigate?) and receiver characteristics (including health literacy and current health status) interact to influence the degree of credibility Internet users assign to the content they find online.

These characteristics influence surface credibility, which determines if, once a user finds a website, they stay on it long enough to find the information they were looking for, and secondly message credibility, which influences whether people feel they can trust the information they find on the website enough to read it. If the website passes these two tests, the user moves to evaluate the

content of the message. Understanding what influences these characteristics and ensuring that appropriate credibility markers are built into health information websites may enable users to be 'leapfrogged' through the model and helped to answer 'yes' to the questions of 'Am I ready to believe this information?' and 'Am I ready to act on it?'.

I wondered how the advent of Web 2.0 and social media might affect the model, however, as it considers the receiver characteristic of a single receiver only, not the receiver characteristics of a social network or group who might receive information collectively. To what extent might a receiver community be able to 'add on' credibility markers that would indicate to other community members whether information was considered credible or not? Many online forums allow users to vote on submitted content, and/or to comment, 'like' or give ratings, providing signposts of approval or disapproval.

This informed my study design by suggesting that it would be useful to test if information that is better received by the community is indeed better quality. If this is upheld, it would suggest that the online community can collectively play an important role in how health information is received by individuals.

In addition, Web 2.0 characteristics encourage constant interaction between the website's operators and its readers, ensuring that content on the site is constantly refreshed and updated, with the frequency of the updating visible to readers through dates and date stamps on content. This shows the receiver that the website is being maintained regularly, that its owners (and community) 'care' about it and that content on it is likely to be up-to-date and in line with any recent discoveries, debates or changes, providing a further credibility marker. Though this was not considered in Wathen and Burkell's original model, is also worth considering in the Web 2.0 context.

## 1.6 Public Health Emergencies of International Concern

When I began my study in September 2013, my aim was to consider the potential – but largely theoretical – use of online discussion forums in future outbreaks of serious infectious disease, particularly those designated as a Public Health Emergency of International Concern (PHEIC) (WHO, 2016a). PHEICs are health events that are formally declared by the Emergency Committee of the World Health Organization, operating under the International Health Regulations 2005. The declaration formally acknowledges a public health crisis that has potential global reach and which requires a coordinated international response. The term (and the operational guidance around what to do when one is declared) was introduced in 2005, following the 2003 outbreak of the previously unknown disease Serious Acute Respiratory Syndrome (SARS) in which 8,098 people became infected in 17 countries (CDC, 2004) – the majority in China and Hong Kong, with significant smaller outbreaks in Canada, Singapore and Taiwan. The outbreak spread from Hong Kong to Canada in just three weeks and led to 774 deaths (Nguyen-Van-Tam & Sellwood, 2013b).

At the start of my study, only one PHEIC had ever been declared – in April 2009, in response to an outbreak of pandemic flu caused by the influenza virus A/H1N1pdm09, commonly called ‘Swine Flu’ as it is thought to have passed to humans from pigs. This PHEIC lasted just over a year, being formally declared over in August 2010. From the early stages of the outbreak, it was apparent that symptoms were relatively mild, with the case fatality rate estimated to have been <1% (Garske et al, 2009; Dawood et al, 2012). In contrast to the pandemic flu outbreaks of the 20<sup>th</sup> Century, which saw significantly increased mortality (>2.5% for the 1918-19 flu pandemic (Taubenberger and Morens, 2006) for example) the disease posed little greater threat to health than normal seasonal flu. An estimated 50-80% of people who

contracted Swine Flu in the UK and displayed some symptoms did not seek a clinical diagnosis (Donaldson et al, 2009) but in April 2009, in the week following the announcement that cases had been identified, the NHS telephone helpline received more than 1,000 calls a day from people who believed they had symptoms, and traffic to the NHS Direct Website (now defunct) increased by 56% (Sturcke, 2009). This suggests that information-seeking did increase, including online, and that resources were challenged because of this.

However, due to the mild symptoms, Swine Flu did not provide a good case study for how information seekers may search out information during a future health emergency with an estimated high mortality rate, as it had not required concerted behavioural changes to contain the disease and prevent further spread, nor had it significantly reduced access to normal healthcare provision. SARS might have provided a better case study, as case fatality rates were as high as 71% in the early stages of the outbreak, later stabilising to 15-17% (Fung and Philip, 2003), but the outbreak largely pre-dates the use of social media and Web 2.0 as we would recognise it, and so was not suitable either.

Without obvious case studies to focus on, my early research design instead focused on the accuracy of health information in online discussion forums, and whether this is high enough for such platforms to be worth pursuing as a possible solution to health-information seeking during pandemics. A short pilot study for this, published in the *Journal of Medical Internet Research* (Cole et al, 2016), and incorporated into Chapter 6 of this thesis, suggests it is, supporting the conclusions of previous studies (Eysenbach et al, 2002; Esquivel et al, 2006; Nölke et al, 2015). The information gained from this pilot study also helped to inform the design of the main study by suggesting that a single comment on a discussion forum is the correct unit of information to assess for quality, rather than the discussion forum 'thread', which contains the initial question or comment and all replies to it. This will be discussed further in Chapter 3, where my methodological approach is discussed in detail.

## 1.7 Ebola Virus Disease: 2014-16 West Africa outbreak

In early 2014, events conspired to change the course and design of my study from theoretical assumptions of what might happen during a PHEIC to observations of what really did happen. Reports of cases of Ebola Virus Disease (EVD)<sup>1</sup> in the West African countries of Guinea, Sierra Leone and Liberia, caused by the strain *Ebolavirus zaire*, began to suggest that a more likely case study was about to emerge.

Ebola Virus Disease (EVD) is a severe viral disease with a case fatality rate of 50-80% and, in early 2014, no known cure or vaccine. From an index case in Guinea in December 2013, the outbreak spread to neighbouring Sierra Leone and Liberia. WHO officially recognised it as Ebola on 25<sup>th</sup> March 2014 and declared a PHEIC on 8<sup>th</sup> August 2014. In September 2014, the U.S. Centers for Disease Control estimated a possible 1.4 million cases in Liberia and Sierra Leone by January 2015 (CDC, 2014) unless interventions were successful and changes in community behaviour could be implemented, as local funeral customs in which large numbers of mourners anointed the still highly contagious body had been identified as a major cause of disease spread (Pandey et al, 2014). The situation in West Africa looked bleak, but did provide an ideal opportunity to study health-seeking behaviour in a *real* PHEIC in *real* time, moving the research out of the realm of the theoretical.

This provided an opportunity for me to ask individuals who were in the middle of the event what information they wanted, and where they looked for it, while being mindful not to distract them from vital work. I conducted semi-structured interviews with 14 employees of NGOs and international

---

<sup>1</sup> At the time, the disease was more commonly known as Ebola Haemorrhagic Fever (EHF). The name was changed during the 2014-16 PHEIC as, while the disease can cause haemorrhage, this occurs in only a small percentage of cases. The name EHF was therefore considered to be confusing, and could prevent milder symptoms from being recognised as Ebola.

companies stationed in Sierra Leone and Liberia. The level of Internet penetration in West Africa at the time of the outbreak was extremely low (only 4.6% of the population had fixed or mobile Internet access in Liberia, 1.7% in Sierra Leone and 1.8% in Guinea (Internet World Statistics, 2014)), and the use of the peer-to-peer networks common in Europe and North America was still nascent. Those interviewed reported health information provision coming mainly through local newspapers and radio shows but the interviews did provide valuable insights into the type of questions people asked, suggesting questions changed depending on the stage of the outbreak and how at risk they personally felt. This finding is consistent with the planning assumptions of the CDC Pre-Event Message Development Project (PEMDP) (Wray and Jupka, 2004; Wray et al, 2008) but contributes to the existing knowledge by indicating exactly what concerns interviewees had, and how these questions related to this specific situation. The full results of this study are presented in Chapter 5 and have also been published in the *Journal of Business Continuity and Emergency Planning* (Cole and Watkins, 2015).

The interviews supported my proposition that online peer-to-peer discussion forums may offer a valuable platform for the exchange of health information during a PHEIC. This gave me confidence that researching the characteristics of such a platform and how one might be configured is valid and worth pursuing. Interviewees placed great value on being able to ask questions directly to individuals, such as on radio phone-in shows where members of the public could express their concerns and discuss specific issues with experts. This suggested that it would be worth exploring how a platform might enable question-and-answer exchanges beyond a static outward push of information.

However, while it may be technically possible to configure a platform capable of providing people with the information they require during a PHEIC, will they use that platform? My experience as a policy researcher at RUSI has

shown that technology is not always accepted by its intended end-users due to human and organisational factors (Cole, 2010) no matter how well it works. While the PHEIC itself provided the context for a perfect case study, I needed to find a peer-to-peer online platform that was being widely used to exchange information on the outbreak.

## **1.8 r/ebola: a case study**

The opportunity to study an actual forum on which information about the Ebola PHEIC was being exchanged was presented by r/ebola, a subsection of the website reddit (www.reddit.com) dedicated specifically to the topic of Ebola. At the height of the Ebola outbreak, the subreddit was receiving more than half a million page views a day (565,280 on 15 October 2014). During October 2014 alone, 339 separate questions relating to Ebola were posted to the forum. Reddit is one of the world's 25 most popular websites (and the 6<sup>th</sup> most popular website in the UK) (Alexa, 2017) – a level of activity that suggests it is considered a credible source of information by a significant number of people. How and why this credibility this is achieved may give insight into potential best practice for such platforms.

Reddit is an 'aggregator' site, which allows its users to collect content from elsewhere, and deposit links to that content into more than 800,000 'subreddits', each dedicated to a specific topic such as r/news, r/starwars, r/diabetes, r/diy, which are set up and run by reddit's registered users. This enables reddit users to direct one another to content from across the Internet they feel is likely to be of interest to the reddit community, and users can vote on submitted content so that other users can see how well it has been received. Users can also comment on content submitted by others in typical message board-style comment trees. One of these subreddits is r/ebola, which was first set up in May 2013 by a regular reddit user who had developed an interest in

Ebola after learning about it in High School. There was little activity on the forum, however, until a year later when two other regular reddit users, who had been discussing the emerging Ebola outbreak in West Africa on the subreddit r/worldnews, decided they would prefer a space for more focussed discussions, and began using r/ebola to share news stories and discuss the outbreak. Their interest, which resulted in them being invited to join the original moderator in managing and administering r/ebola, saw it grow to 565,280 page views a day and 14,224 subscribers at the height of the outbreak in mid-October 2014.

Traffic and subscriptions amongst its users (more than 50% of whom are based in the U.S.) peaked when the first handful of cases reached U.S. soil, suggesting that reddit may be the type of platform to which significant numbers of health information seekers turn for information during the critical early stage of a local outbreak, during which non-medical interventions can have a strong impact (WHO, 2008; Funk et al, 2009; Nyugen-van-Tam & Sellwood, 2013a). This made it a perfect case study for my main research. As the outbreak and the opportunity offered by it developed, the study design changed with it, resulting in several small studies rather than a single large one, which were then synthesised and compared against one another.

By this stage, I knew from the pilot study that the quality of information on reddit was generally high (Cole et al, 2016) and from the interviews with NGO workers in West Africa, I had an idea of what questions people were likely to ask. r/ebola allowed me to now test this against a PHEIC-specific website, and to see if the discussions matched those that the NGO workers had indicated they would find useful. The content of the questions asked provided information on what concerns users had. The answers they received, which were voted on by the r/ebola community, could be evaluated for quality by asking doctors to assess them. These assessments could then be compared

with the evaluations given by the reddit community to see if reddit users were able to recognise ‘good’ and ‘bad’ information.

For the main phase of my study (described in more detail in Chapter 3), I observed activity on the r/ebola forum and interviewed its moderators. I asked doctors to rate the answers provided – as I had in the pilot study, but this time with specific reference to individual comments rather than entire threads – and cross-referenced the ratings given by 27 doctors with the votes those answers received from the reddit community. If the answers the doctors consider to be ‘good’ are also the ones upvoted by the community, this will uphold the idea that the community directs its members towards information they *should* trust. Observing the forum in real time also alerted me to challenges it was facing and enabled me to assess how the forum operators and users dealt with this.

## **1.9 Gaps in the current understanding**

This study identifies and seeks to fill gaps in the current understanding of health information forums and their potential use in a PHEIC. Firstly, it will explore whether online discussion forums have certain characteristics that may benefit health information seeking online and how reddit builds user confidence in its platform (whilst acknowledging that disadvantages may also become apparent). The few previous studies there have been into the quality of information in such forums show that incorrect information tends to be swiftly removed or corrected by other users (Esquivel et al, 2006; Mursch and Behnke-Mursch, 2003) suggesting that the discussion forum format is well-suited to answering health questions accurately and offers more potential than is currently being acknowledged or exploited. Others have questioned this, however, arguing that as health forums are mostly used by ‘expert’ patients (Wilson, 2001) with extensive experience of living day-to-day with their conditions, the findings may not be easily generalizable to all health

conditions (Dhatariya, 2006). If true, this would present a real challenge to the use of such forums during sudden outbreaks of unfamiliar disease, so it is important to explore the extent to which the 'expert patient' is a factor in quality control, and how this compares with other factors.

Secondly, while studies have focussed on the ability of discussion forums to provide an insight into patients' views (Benton et al, 2011), none have looked at this in the specific context of a public health emergency or have suggested how websites might best be configured to provide the information most required during such events. Websites such as reddit enable users to not only ask a question, but to have that question answered in a way that is specific to their needs. The answer(s) they are provided with can be corrected, elaborated on or countered by subsequent posts. Some previous studies have identified this as a potential or actual benefit of discussion forums (Hoch et al, 1999; and Feenberg et al, 1996; Mursch and Behnke-Mursch, 2003; Esquivel et al, 2006) but none have looked at this in the specific context of a public health emergency.

By focussing on reddit and r/ebola, rather than the many other websites that were offering health information during the Ebola outbreak, my study will overcome one of the main weakness of the existing literature on health information online (and in fact, on information online in general), which is the tendency to approach 'the Internet' as if it is a homogenous environment where users trust or mistrust every website equally, at least until trust is added on by externally awarded accreditation seals or source authority can be proven (Akerkar and Bichile, 2004; Lawrentschuk et al, 2012). I will determine whether some websites' characteristics can signpost good quality information, making them inherently more conducive to the transfer of high quality information, and more suited to health-information seeking, than others.

## **1.10 Contribution to knowledge**

The contribution to knowledge I aim to make with this thesis is a greater understanding of how, during a PHEIC, it might be possible to configure peer-to-peer platforms to help health-information seekers locate high quality information and navigate the potential dangers. I will determine if users are able to do this more easily and safely than the healthcare professionals I surveyed currently give them credit for, guided by the collective intelligence of the online communities and the trust signposts embedded in the platforms. If this is upheld, it will suggest that greater utilisation of discussion forums is possible and that they could be used during a PHEIC to relieve pressure on stretched or overwhelmed professional health providers.

Using r/ebola as a case study, I will provide deeper understanding to academics, clinical staff and public health officers of the structure and technology of a widely used PHEIC forum, so that its potential usefulness – or the usefulness of a platform like it – in a public health emergency can be assessed. The usefulness of such a platform depends on how accurate the information it contains is shown to be; the degree to which people do, and can be encouraged to, trust it; and how the collaborative space it provides enables more appropriate information to emerge than might be provided by other types of platforms. The lessons identified from r/ebola will therefore provide strong foundations for configuring or modifying other platforms in future.

During my data gathering, health professionals appeared wary of peer-to-peer information sharing platforms. They assumed information found there would be poor quality and that receivers, whose health literacy they doubted, would be unable to recognise this or to make sensible decisions in the face of it. As a result, they see little potential for such platforms. The precise reasons why they are so sceptical would be interesting to know, but have been largely

outside the scope of this study to determine, other than to acknowledge that such prejudice exists. It would be a valuable subject for further research.

### 1.11 Scope and structure of the study

The scope of this thesis spans the fields of human computer interaction (HCI), Online Health Information (OHI) and Public Health Communication. At its foundation is a desire to improve public health communication strategies by understanding the appropriateness of online peer-to-peer platforms to the context of a public health emergency.

Over the following chapters, this thesis will answer the main research question and five sub-research questions:

SRQ	TOPIC	ADDRESSED IN
Main RQ	In what way(s) do the characteristics of online discussion forums facilitate or hinder health information seeking online?	Chapter 4, 5, 6, 7, 8
SRQ-1	What information do people search for during a public health emergency?	Chapter 5
SRQ-2	What technology affordances of online discussion forums might help users to trust the information found there?	Chapter 6
SRQ-3	Is the quality of information in online health discussions of sufficiently high quality to be of value to health information seekers?	Chapter 6
SRQ-4	What characteristics of discussion forums help to maintain or compromise information quality?	Chapter 7
SRQ-5	How might this influence the configuration and/or utilisation of online discussion forums during a future Public Health Emergency of International Concern?	Chapter 8, 9

TABLE 1: *The main research question and sub-research questions addressed in this thesis.*

The structure of the thesis is set out on the following pages.

Following the introduction in **Chapter 1**, **Chapter 2** sets out the Theoretical Framework for this research, how I used it to shape the choices I made and how these influenced my methodological approach. It explains why I drew from literatures on affordances (Gibson, 1979; Norman, 1986, 1999); Human-Computer Interaction (Hiltz and Turoff, 1978; Davis, 1989); and Crowd Wisdom and Collective Intelligence (Galton, 1907; Surowiecki, 2005; Lévy, 2007, 2010; Malone, 2008, 2015), and how I applied this to the exploration of online spaces where people sought out and shared health information during a serious disease outbreak. I consider not only the health information itself, but also how that information is managed by the technical architecture of the system and the behaviour and processes of the online communities that contribute to and draw from it. Such platforms may need to encourage collaboration and group working, for which I explore how groups receive information; store and retrieve that information and come to decisions. I anchor this approach on the Theory of Collective Intelligence (Pierre Lévy, 1995, 1997) to which I will return throughout the thesis to orient the research findings, and to assess whether online platforms enable Collective Intelligence to emerge in shared online spaces. As all spaces exert influences on the choices that people make in them (Thaler and Sunstein, 2008) and better designed online spaces are more likely to influence users' choices in the desired direction (Munson et al, 2013), doctors, healthcare policymakers and those affected by a PHEIC, as well as academics, will benefit from a better and more nuanced understanding of the online environment in this context.

**Chapter 3** describes the methodology used and explains its rationale.

**Chapter 4** takes a detailed look at the case study platform r/ebola, and attempts to determine which of its characteristics may help or hinder the emergence of collectively intelligent health information.

Having determined how r/ebola *may* provide a mechanism through which good quality health information *may* emerge, **Chapter 5** will determine which questions health information seekers are likely to want answered during a PHEIC, and **Chapter 6** will determine the extent to which r/ebola *did* answer health questions accurately, through two studies that asked qualified medical practitioners to rate the quality of health information found across three reddit health forums, and then on r/ebola specifically.

**Chapter 7** will examine a platform characteristic identified in Chapters 5 and 6 as most efficacious in quality assurance – moderation – and will identify the precise mechanisms by which the process of moderation influences information quality.

**Chapter 8** will bring these themes together, and review how they were able to influence and enable the rapid scaling up of the r/ebola subreddit, which grew from just 509 subscribers on the day the Ebola PHEIC was officially declared, to more than 14,000 subscribers three months later. I will identify what made it so successful but also highlight what challenges it faced.

**Chapter 9** will then bring together findings from across Chapters 4-8 to suggest how to configure peer-to-peer information-sharing platforms in future PHEICs, including the human elements of the system, as well as the technical components.

## **1.12 Limitations of the study**

My research findings are limited by the specific case study chosen and the relatively small sample sizes. While my qualitative data samples are small, however, they are high quality – I interviewed 14 people who were witnessing the Ebola outbreak first hand, and five who had moderated forums dedicated

to discussing it as it unfolded. Nonetheless, it cannot be assumed that findings from this thesis can be generalised to all populations, in all types of health emergency, nor even to U.S.-based users of reddit during a future outbreak of *Ebolavirus zaire*. The emergence, characteristics and use of a subreddit, or any other online forum, dedicated to a future PHEIC may or may not mirror what was observed on r/ebola. What is inferred from this study should be taken as a suggestion, not a definitive design architecture, for what might prove useful in future. Results gained from the study of r/ebola need to be checked against other forums in other PHEICs.

### **1.13 Conclusions**

This study therefore aims to explore whether online peer-to-peer discussion forums offer a valuable platform for the exchange of health information during a PHEIC. I will explore whether such forums may be particularly useful during a health emergency involving a novel or rare disease about which little is known, and for which no large-scale outbreaks have previously been recorded, resulting in a lack of certainty about how the outbreak might develop and limited expertise amongst the scientific, medical and policymaking communities tasked with bringing the outbreak under control. How might such systems enable members of the at-risk community to share information and experiences, ask questions and (where necessary) come to a group decision on what action should or should not be taken? Through this, I will explore whether the public might be able to play a key part in the response to the outbreak, freeing up limited professional healthcare resources to focus where they are most needed.



## 2. THEORETICAL FRAMEWORK AND LITERATURE REVIEW

---

### 2.1 Introduction

The aim of this thesis is to explore whether online discussion forums provide useful spaces across which health information can be exchanged, particularly during times when access to doctors and professional healthcare organisations might be constrained. To answer my main research question – *In what way(s) do the characteristics of online discussion forums facilitate or hinder health information seeking online?* – I need to explore two ideas. First, do online discussion forums have certain characteristics that might help or hinder health information seeking? Second, if they do, how do these characteristics influence the process of information seeking to help or hinder it?

I approach the first idea – that online discussion forums may have certain characteristics that help or hinder health information seeking – by considering the concept of technology affordances (Gibson, 1979; Norman, 1988, 1999), in which technology affords the user an agency they may not have otherwise, or not have to same extent. Here and throughout the thesis, I will refer to the concept of technology affordances to describe the characteristics I observe from my study of reddit and other online platforms.

I approach the second suggestion – that these characteristics do indeed help or hinder online health information seeking – using the concept of Collective Intelligence (Hiltz and Turoff, 1978; Lévy, 1995, 1997, 2010; Malone, 2008), which considers, to quote the main research question of the Center for Collective Intelligence at Massachusetts Institute of Technology (MIT): “How can people and computers be connected so that – collectively – they act more intelligently than any person, group or computer has ever done before?”. The key difference (or technology affordance) between discussion forums and

other types of online platform is that they offer opportunities to discuss and debate the information found there. Can a ‘Collective Intelligence’, greater than the sum of its parts, emerge from this collective action?

During a health emergency initiated by a novel disease about which little is known, a discussion forum may fulfil a similar function to the Delphi Method, a system for refining collective judgements in situations where knowledge is fragmented between a group of partial experts (Dalkey, 1969; Ghamari-Tabrizi, 2000). The communication opportunities offered by discussion forums provide us with the ability to make decisions not only individually, but also collectively. Bonabeau (2009), writing from the field of organisational and management theory, called this ‘Decisions 2.0’.

Discussion forums might also be a proxy for the focus groups used in Robert Chambers’ Rapid Rural Appraisal (Chambers, 1981, 1992), pooling the inherent knowledge of the community to achieve a ‘proportionate accuracy’ that, while it might not be perfect, is good enough for the context in which it is required.

My research design draws on the disparate theoretical frameworks presented above to converge on the suggestion that online health forums not only enable discussion and collective decision-making but may also result in a form of Collective Intelligence that is particularly appropriate when resources are constrained. My conceptual framework asks what technology affordances of an online platform – and of reddit as a case study – are most conducive to the emergence of Collective Intelligence in such situations? In this chapter, I will introduce these concepts; provide a review of key literature, including criticisms of the main theoretical approaches; and cite examples of their applications. Throughout this and later chapters, what is known and learned

about online health information forums in general, and about reddit and r/ebola particularly, will be considered within this theoretical framework.

## **2.2 Technology affordances**

To prove or disprove the idea that online discussion forums have characteristics that might help or hinder health information seeking, I first draw on the Technology Acceptance Model proposed by Fred Davis (Davis, 1986) and User-Centered System Design (Norman 1986), popularized in Don Norman's 1998 book *The Psychology of Everyday Things* (later renamed the *Design of Everyday Things*) from the field of human computer interaction (HCI). These approaches consider how the design of technology can influence and affect how people interact with and use, or do not use, that technology. This will indicate what affordances reddit offers and what actions they enable.

The term 'affordances' was coined by the psychologist James J. Gibson (1904-1979) to describe the properties or qualities of an object – the ability, or 'agency' it affords the user (Gibson, 1979) – where agency is defined as the capacity of the actor to act in a given environment. For example, a handle 'affords' grasping and holding; this in turn enables 'carrying'. Gibson's work related to environmental and evolutionary biology. He theorized that during human evolution, mankind has continually modified the world around us to change what the environment affords us, thus increasing our agency, while also stressing that the world of material products is not distinct from the mental products we have used to shape it. We have not created a new space, but have modified the existing one to suit our ends. Salomon (1993, p51) describes affordance as referring to the perceived properties of a thing, which determine how it could possibly be used, as well as the actual properties. The concept of technology affordances has been criticised for failing to acknowledge that affordances can be negative as well as positive (Kaptelinin and Nardi, 2012) and that they may be visible, hidden or false (Gaver, 1991,

1992). For an affordance to be useful, it must be obvious to the user. Even then, difficulties in using it may lead to it being resisted, so that it should be considered a possibility for action, rather than a given (McGrenere and Ho, 2000). Objects can also be ‘disobedient’ (Grindon and Flood, 2014), appropriated by social movements and actors for purposes well beyond the original intention of their designer, for example plastic bottles can be used to make tear-gas masks for use during riots. Technology affordances are dependent on social and cultural as well as mechanical factors and so should always be considered in context (Kaptelinin and Nardi, 2012), including whether the affordances benefit individual or collective action. This will be borne in mind throughout my research, not least because reddit is a prime example of how affordances can ‘shift’ beyond original intentions. Conceived as a platform through which to order food (Kersey, 2012), computer entrepreneur Aaron Swartz, one of the developers of Rich Site Summary (RSS) software, saw its potential as a news aggregator on which users could rate the content. He became one of its early backers, in 2005. When Condé Nast, an international publishing house, bought in a year later, its interest was reportedly in reddit’s ordering algorithms (Arrington, 2006; Wired, 2006), which it wanted to use on its make-up review site [www.lipstick.com](http://www.lipstick.com). A platform’s flexibility to conform to different user contexts and requirements may be a key characteristic, particularly during events that are fluid, unprecedented or unpredictable, as a serious disease outbreak is likely to be.

‘Affordances’ was appropriated for use within the field of HCI by Don Norman (1988, 1999) building on earlier work in which he described the challenges of ‘Cognitive Engineering’ (Norman, 1986). Norman sought to understand how people use technology, so that it can be designed in a way that enables potential users to make a quick cost-benefit analysis of its value. According to Norman (1988), affordances provide clues to the operation of something; if a doorknob is well-designed, which way the user should turn it

should be immediately obvious: any confusion suggests that its design could be improved. This should be as applicable to computer systems as to any other form of technology or object. A well-designed health forum should enable a user to easily find the information they require, and to assess whether the information they are viewing is valuable and appropriate to their needs. This might be seen to 'hold their hand' through Wathen & Burkell's model for assessing credibility (Wathen & Burkell 2002), providing positive indicators of surface credibility and message credibility, and helping them to evaluate the message content.

Fundamental to people's use of technology, Norman considered, is that using the technology needs to be "pleasant, even fun" as well as efficient, as the user is likely to have a psychological goal, directly related to their needs and concerns, which exists in parallel to the physical system. He calls this the psychological goal: the state the person wishes to achieve and which has value for them. The psychological goal of a reddit user posting a question on r/ebola is probably not 'to use reddit', but to receive an answer to their question, which they hope will be provided by its intended readers. Reddit affords them a place where they can ask their question, and access to people who may be able to answer it. Equally important are affordances that enable them to decide which of the millions of reddit users might be best able to provide an answer and which answer(s), if more than one is returned, is the best or most accurate.

The reddit user's intention to act, and so to achieve their goal, may be impeded if the technology looks too difficult to use, if they think they may look stupid trying, if they think the answer they receive may be of poor quality, and so on. To ensure technology is acceptable to the user and will be not be rejected by them, the technology affordances of the system must not place barriers in the path of the psychological goal. Norman theorized an 'execution gulf' and an 'evaluation gulf' between the physical system and the psychological goal. If one or both gulfs is not bridged, the technology is

unlikely to be accepted by the user and will not afford the user agency. The challenge is to ensure that the conceptual model of the designer during the development of the system matches the conceptual model of the intended end-user: to practice what Norman called User-Centered Design (UCD). This starts with the needs of the user, recognizing that the aim of the system is to serve those needs, not to have the user operate a specific technology.

Online communities have been identified as particularly conducive to and representative of UCD (Preece et al, 2004) as over time the interactions of the user community lead to the technology adapting and evolving to meet the needs of that specific community. A criticism of UCD (Abrams et al, 2004) is that it takes more time and involves more design team members and – even when the luxury of time is available – it can cost more. Its dependence on there being time available for it to happen may be especially problematic in a health emergency that arises rapidly.

A second criticism – that it can end up providing a system that is very specific for one context and one group of users – may be less of a challenge here, however, as a system may only need to be relevant for the duration of one disease outbreak.

The technology affordances ICT systems may offer have been further explored by Conole and Dyke (2004), who developed a taxonomy for different types of affordance, including speed, collaboration and communication, that may help practitioners to understand the technology; identify limitations (constraints); enable discussion that can lead to critique and refinement; and work as a checklist of the cost-benefit analysis. A technology affordance of Skype, for example, is that it enables visual and audio communication across the Internet; a constraint is that it requires a computer screen and a reasonably good Internet connection. Reddit will have similar constraints. When considering how it is used, and might be used in future, I need to bear in mind

its likely constraints as well as its affordances as these might also influence users and their interactions with the technology available.

### ***The Technology Acceptance Model***

To understand how reddit might be used, and what advantages or disadvantages it might have over other platforms, I consider it in the context of the Technology Acceptance Model, or TAM (Davis, 1986 and 1989; Davis et al, 1989), which seeks to explain why a user may accept or reject information technology. This study will provide unique insight into the way in which reddit maps onto TAM and which of its features enable or constrain users' attempts to source health information during a health emergency.

Potential gains that might be provided by online health discussion forums and particularly those that are hosted by reddit could be obstructed by users' unwillingness to accept and use available systems. Two key factors influence user acceptance. First, people tend to use, or not use, technology to the extent that they believe it will help them to perform a task (e.g. finding health information) better than they are able to without it: its *perceived usefulness* (PU). Second, and potentially conflicting with the first factor, is whether the user thinks the system is so hard to use that any performance benefits are outweighed by the effort of using it: the *perceived ease of use* (PEOU). This is a straightforward example of the cost-benefit paradigm from behavioural decision theory (Beach and Mitchell, 1978), and can also be broadly equated to Norman's execution/evaluation nexus, as well as to the Theory of Reasoned Action (Fishbein, 1979) and the Theory of Planned Behaviour (Ajzen, 1985 and 1991), which seek to explain and predict behaviour in specific situations. Humans consider the implications of their behaviour before deciding whether to follow a course of action, influenced by Perceived Behavioural Control (PBC), which may constrain their freedom to act. TAM suggests PEOU and PU are the two most important PBCs in determining system use: throughout

this thesis it will therefore be important to address how the technology affordances of the discussion forums on reddit may influence their PU and PEOU to health information seekers.

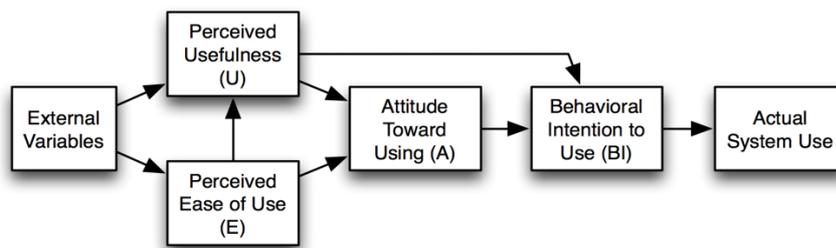


FIG 2: *Technology Acceptance Model (Davis, Bagozzi & Warshaw, 1989)*

The TAM is not without criticism. It was developed as an indicator for the likely uptake of work-related technology systems, which people (at the time, mostly computer scientists and IT professionals) had an option to adopt or not. As technology has developed and become ubiquitous, users may no longer have a realistic choice of whether to use it, and may have to overcome difficulties with perceived ease of use (Brown et al, 2002; Hwang et al, 2016). However, as reddit is an ‘opt-in’ system, which users make a conscious choice to use, and more than 360 million people worldwide have made this decision (in June 2017, Alexa.com ranked it as the 4<sup>th</sup> most popular website in the US, the 6<sup>th</sup> in the UK and 9<sup>th</sup> in the world), I will assume that its PU and PEOU are both relatively high. Examining reddit to understand why might help to inform the design of future systems.

A second criticism of TAM is that the model relates to a single point in time, and does not account for how the PU/PEOU relationship may develop and change over time, particularly in settings where the user is obliged to use the technology. This is particularly interesting to consider in the context of reddit.

As it has grown and evolved over more than a decade, with its user community modifying aspects of the platform technology throughout that period, a new subreddit dedicated to a disease outbreak may be able to ‘carry over’ PEOU and PU from one or more of the existing subreddits, and be more able to anticipate the PU/PEOU relationship of its potential user community than an entirely new forum. I will consider this in Chapter 8, when I examine how the r/ebola subreddit was set up and developed.

Criticizing TAM became popular in the mid-2000s, but more recent studies (Hwang et al 2016) have argued that as modifications are made to the original, the developed versions seem to fall back to an even closer allegiance to the theories on which the original was based. This arguably validates it, though its inability to suggest an implementation strategy for new technologies remains a common criticism (Venkatesh et al, 2003) and more research on implementation is required (Venkatesh and Bala, 2008).

Later modifications of the TAM have aimed to improve it, leading to TAM2, The Unified Theory Acceptance and Use of Technology (Venkatesh and Davis, 2000; Venkatesh et al, 2003) and later TAM 3 (Venkatesh and Bala, 2008). These have put more emphasis on the role of external variables than was in the original model, including user experience, context relevance and voluntariness of use, just as considerations of social and cultural contexts have been added to technology affordances in general. This has sought to address some of the main criticisms and to modify it accordingly. As a result, the TAM still has many supporters and I argue that it is a solid base from which to consider the usefulness of reddit, while reminding me to keep the influence of external variables and social contexts in mind. TAM has been tested on many computer applications including word processing tools, computer banking systems, email and online shopping systems, but a systematic review of the TAM literature (Hwang et al, 2016) did not cite any examples of the TAM

being applied to online discussion forums. My intention to not only consider how reddit currently reflects the TAM, but also how a health emergency might influence the PU and PEOU of a system, requiring that relationship to change and evolve over the course of an outbreak, therefore contributes to this existing body of knowledge. While it is outside the scope of this study to explore, I am mindful that the stressful context of a PHEIC is unlikely to be a good time to expect those involved to adopt new technology. Whether the technology has already been accepted, or needs to be accepted while the PHEIC is in progress, may be an important consideration.

I explore the technology affordances of reddit in depth throughout Chapters 4, 5, 6 and 8, as I consider how reddit provides access to health information; how accurate that information is; whether users are guided towards higher quality information and away from poorer quality; and whether affordances that enable this were applied on r/ebola to serve the needs of the user community during a Public Health Emergency of International Concern. In doing this, I will keep in mind how the concept of affordances has developed, and be mindful of the influence of social and cultural as well as mechanical factors (McGrenere and Ho, 2000; Grindon and Flood, 2014), and factors that may benefit collective action (Kaptelinin and Nardi, 2012). This will help me to consider which affordances may encourage Collective Intelligence to emerge.

### **2.3 Collective Intelligence**

*“The extent to which [society will benefit in the future] will be based upon collective intelligence. This is to society what brainpower is to the individual.”*

– Lester Frank Ward (Ward, 1906, p39)

A key aim of this thesis is to explore whether the connectivity reddit offers, to a health information-seeking community and to a wide collection of health

information, provides measurable benefits to a reddit user. To understand this, I will first consider how Collective Intelligence emerges in human groups; then look at the early computer mediated communication (CMC) systems to which the phrase was first applied in computer science by Starr Roxanne Hiltz and Murray Turoff in their book *Network Nation: Human Communication Via Computer* (Hiltz and Turoff, 1978). I will reflect on the new types of group interaction CMC enabled, and then turn to considerations of how this human-computer network may be particularly conducive to the emergence of Collective Intelligence.

The concept of Collective Intelligence far predates its potential application to computer networks. The phrase ‘two heads are better than one’ is an old English proverb which appears in print as early as the mid-16<sup>th</sup> century (Heywood, 1546) and has been remarkably enduring. In 1906, Sociologist Lester Frank Ward wrote, “The extent to which [society will benefit] will be based upon collective intelligence. This is to society what brainpower is to the individual” (Ward, 1906, p39). Ward believed that individual geniuses could do nothing without a social structure that enables them to emerge, supports them and allows them to thrive. He wrote that society acts collectively: only by working together through an enabling mechanism can individual members of society increase the intelligence of that society. At its heart, his message is that a group is more intelligent than its individual constituents. If this is true, it suggests that providing a space in which a group can come together to discuss health information might help to increase the intelligence, or health literacy, of individual group members.

The same year as Ward wrote about collective intelligence, polymath Francis Galton visited a country fair where he observed a competition to guess the weight of an ox. Eight hundred people entered the contest, following which Galton did a statistical analysis on the answers provided and found that the

median of all guesses (1,207lb) was extremely close to the actual weight (1,198lbs, a 0.8 percent deviation) and more accurate than the guesses made by almost all experts. Furthermore, few of the estimates were wildly inaccurate – most clustered close to the accurate median. When he published his findings in *Nature* as *Vox Populi* (Galton, 1907), the idea of crowd wisdom – that the crowd working together can be somehow better than the individuals that comprise it – began to gain popularity and academic acceptance.

### ***The Wisdom of Crowds***

Galton's example is one of many used nearly a century later by financial analyst James Surowiecki in *The Wisdom of Crowds* (Surowiecki, 2005). He observed numerous situations in which a group of individuals appeared able to find solutions to very sophisticated problems, such as the stock market correctly identifying which manufacturer's component was at fault soon after the spaceshuttle *Challenger* exploded: their stock fell more quickly than that of the other three companies who had supplied components to the spacecraft. Surowiecki provided similar examples from prediction markets and sports betting, and proposed that in the right circumstances, a group can be "smarter than its smartest member". The correct identification of the shuttle crash has been attributed to 'crowd' awareness of problems that had been identified over the previous year (Maloney and Mulherin, 2003), which allowed enough people to make an educated guess as to what had happened that the 'crowd' could reach a correct conclusion; though not, it is worth noting, to have predicted the likelihood of such an event happening and to have acted in advance to prevent it. Nor is there evidence that wise crowds have been able to predict and deflect other significant events such as the 2007/8 financial crisis.

Furthermore, while social interaction can enable individual knowledge to be 'shared, corrected, opened, processed, enriched and evaluated' (Schuurman et

al, 2012, p53), critics point out that once interaction takes place, an individual's answer can be biased by social processes such as groupthink (Janis, 1972; Lorenz et al, 2011), in which a desire to conform can sway judgement, and cognitive bias (Haselton et al, 2005), in which preconceived ideas can lead to irrational judgements and impaired decision-making. These lower the overall intelligence of the group and potentially enable conspiracy theories to take hold (Hill and O'Hara, 2006).

Surowiecki nonetheless suggests that advantages of crowd wisdom include cognition (especially 'market judgement', which he considered to be potentially more accurate than detailed academic study); coordination (influenced by common understanding); and cooperation (which can result in a network of trust). Critics (Schuurman et al, 2012; Lanier, 2006, 2010) suggest that while the theory is applicable to objective, measurable information – such as guessing the weight of an object as Galton did, or guessing which out of the four manufacturers of space shuttle components was most likely responsible for the one that had failed – it may be less appropriate for innovation and ideas.

In most of Surowiecki's examples, however, there is none of the coordination or discussion between the members of the crowds that online forums can facilitate; instead, several individuals, acting alone, contribute their answer to a given problem and the answer is calculated from the mean or median. The contributions of the crowd are observed from outside of it and the final decision is made without crowd input. In fact, he considers that for the group collectively to be more accurate than a single individual, it is essential for each member to be able to make their contribution independently, without being swayed by others around them. There must be an unbiased mechanism – he called this a "clever mechanism to turn the individual ideas into a collective decision" – in place to collate the diverse opinions that may emerge without

biasing the individuals making them. Reddit certainly provides a collection mechanism – the ability to post opinions, ideas and answers on a platform that aggregates them – but whether it does this in a way that may bias the views of others is debatable. Users vote on reddit’s content independently, but they can see how that content has been voted on before. There is conflicting evidence of what influence, if any, this has on how votes are cast. Weninger et al (2015) suggest that it does, particularly in the case of upvoting, while others have suggested that it does not (Stoddard, 2015b). This discrepancy will be kept in mind as I investigate how reddit collects and presents information to its users. Surowiecki argues that while disadvantages of crowd wisdom can and should be highlighted, it is often the mechanism for collection that is flawed, not the process itself. This opens the possibility that some characteristics of online platforms may act as a ‘clever mechanism’ to enable or impede the emergence of (more) accurate health information from a ‘wise crowd’ of discussion forum users – if they can guard against biasing the final decisions.

I also acknowledge that the concept of groups of humans acting together in intelligent ways is not unique to online communities. Professor Thomas Malone, founder and current Director of the Center for Collective Intelligence at MIT Sloan offers families, countries and companies as pre-Internet examples of intelligently networked groups (Malone, 2006).

As groups can influence one another negatively as well as positively, it may help to think of collective ‘intelligence’ in the military sense of intelligence: as processed information, which is not necessarily accurate, correct or true, and which needs additional signposts (such as trust in the source, echoing Wathen & Burkell’s model), to verify its credibility and value. Bearing this military terminology in mind may suggest ways to mitigate any such negative effects, not least because the military is where The Delphi Method, one of the earliest, most effective and still most widely used, methods for harnessing collective intelligence was developed – four decades before the invention of the Internet.

### ***The Delphi Method***

The Delphi Method was developed during the 1950s by the American RAND Corporation (a U.S.-based global policy think tank, originally formed to offer analysis to the U.S. military). It is a system for eliciting and refining group judgements for situations where exact knowledge is not available (Dalkey, 1969) and the 'best guess' of a group in which different areas of expertise are represented may provide the best option (Ghamari-Tabrizi, 2000). It aims to minimize the biasing effects of dominant individuals and of group pressure towards conformity, thus providing a guard against groupthink and cognitive bias. In the Delphi Method, a question is presented to a group of experts and each one answers it independently. The replies are then circulated amongst the group and considered collectively (with each participant unaware of who made each contribution). Each contributor is then able to reconsider or amend their response in light of the comments from others. Participants are generally expert scientists and academics, but one can also apply its intention – to bring together several possibly conflicting views so that a consensus can be found, or to collect partial views and to mould them together into a greater whole – as a form of message board without the Internet.

RAND used it during the Cold War as a tool to 'imagine the future' of a potential nuclear strike, an event of which neither military strategists nor academics had real experience. During the 1960s and 70s it was extensively used by the U.S. Office of Emergency Preparedness during times of risk and uncertainty when situations were constantly changing, including an online version EMISARI (Emergency Management Information System), which was in use until 1986 (Turoff, 2002). It has been particularly popular in situations where there is no historical data or when ethical and social dilemmas dominate economic and technical ones (Rowe et al, 1991), though critics point out that a group of experts, each knowledgeable about one aspect of a complex system,

does not necessarily comprise expertise about the total system (Lindstone and Turoff, 1975 and 2011). Where expertise about the total system is not readily available, however, such an approach may provide the next best option.

A systematic review of Delphi Method application (Gupta and Clarke, 1996), found that healthcare is one of the most popular areas in which it has been used. It may have particular value during health emergencies involving previously unknown and newly emergent diseases, when the medical and scientific community may not have all the answers or hold all of the expertise: during the 2014-2016 Ebola Outbreak, expertise from anthropology, public health communication, education theory, geopolitics and behavioural psychology were identified as being as important to the response as medicine, epidemiology and virology (Abramowitz, 2014; Pandey et al, 2014; Leach, 2015). Creating spaces where partial experts can come together and discuss options is likely to provide at least some value.

#### **2.4 How good is good enough?**

On reddit, the constituent members of the crowd may come to discussions with very little, or extremely fragmented, information. Even in non-expert settings, however, the process of pooling and discussing information seems able to improve the quality of the answers given by single participants. It is worth reflecting on the Rapid Rural Appraisal (RRA) methodology developed by Robert Chambers (Chambers, 1981). In this, he introduced the concept of 'proportionate accuracy', where recognising the *degree* [my emphasis] of accuracy required becomes more important than achieving absolute accuracy. Chambers' work focused on developing rural communities, where decision makers need information that is relevant, timely, accurate and usable but, where time is constrained, choices need to be made that maximize the likelihood of the other three criteria being achieved. If the intrinsic knowledge

held by a group can be quickly consulted, and may be nearly as good as a lengthy, more meticulous exploration, it may well be accurate *enough*. Chambers saw group interviews as having several advantages, including access to a larger body of knowledge, mutual checking, and the ability to cover a wider subject matter than can be covered with one respondent. Gordon (1979) observed a self-correcting mechanism within group interviews where members cross-check what they think they know with others.

Translated to the online sphere, *Nature's* study of Wikipedia vs *Encyclopaedia Britannica* (Giles, 2005), found that the peer-developed Wikipedia is very nearly as accurate as the professionally edited *Encyclopaedia Britannica*. This has been upheld by subsequent more rigorous academic research (Casebourne et al, 2012), suggesting that when Wikipedia founder Jimmy Wales made a conscious decision to risk sacrificing accuracy for time in his split from Wikipedia's predecessor Nupedia (Sanger, 2005) it was a risk worth taking. Chambers' approach and methods for learning about a situation from people who are directly experiencing it, might be transferable to a community of online health seekers during a pandemic, particularly if the online community holds deep and tacit knowledge about itself, including locally developed solutions, which can be shared with outsiders – the health-information seeker taking the place of Chambers' academic researcher.

An important point Chambers identifies in Participatory Rural Appraisal (PRA) – a later development of RRA – is, “not to wait, but to start, stumble, self-correct and then share” (Chambers, 1992, p22): knowledge can be spread through the sharing of experience and mutual learning. In group interviews, participants fill in gaps left by others and add or correct detail; they have an overlapping spread of knowledge which covers a wider field and cross-checks the information provided by individuals. When, in Chapter 6, I consider the accuracy of the health information presented by the peer-to-peer networks of

reddit, often in the absence of health professionals, I will keep the concept of proportionate accuracy in mind.

The concepts behind Chambers' approach have also informed the 'agile design' of systems (Dearden and Rizvi, 2008), in which participants contribute knowledge of how best to develop and configure a system over which knowledge is shared and collated, as well as the content itself, echoing the principles of User-Centered Design. Letting the user community define the technical characteristics of the platform it uses can be useful as different characteristics may benefit different contexts (Maloney-Krichmar and Preece (2005). A valuable contribution to knowledge made by this thesis is that my understanding of which characteristics are particularly attractive to the user community in the context of a health emergency is developed not only by my observation of the system but also from the interviews conducted with the reddit moderators who used it.

### ***Maximization and the Act of Choice***

Reddit also invites reflection on the idea of 'maximised choice' developed by Amartya Sen. In 'Maximization and the Act of Choice' (Sen, 1997), he explored the concept of *decision inescapability*, in which a decision must be made even when the conditions under which the ideal (or optimal) decision can be made are not met. As he describes (Sen, 1997, p746):

*“[T]he importance of the act of choice also lies in its inescapability or urgency [...] if there is no escape from choosing, a choice decision will have to be made even with incompleteness in ranking.”*

Sen suggests that under such circumstances, the individual may value not only the alternative they eventually choose, but the fact that they have had the

opportunity to choose and have made their own choice. Where no optimal choice is available but a choice must be taken, the ‘maximal’ alternative becomes the best option.

However, making such a choice is problematic if the person making it has a limited understanding of how sub-optimal any one choice is when measured in isolation; against the optimal choice; or against other sub-optimal choices. If the optimal health-information seeking choice of consulting a doctor cannot be made, how does one assess whether seeking information from many posters on an Internet health discussion forum or from a single source such as a blogger or a static webpage (and if so which blogger or static webpage), is the maximal choice of the options available?

This suggests that a framework is needed to help understand what conditions influence the likely maximization, or not, of Internet sources as the health information provider of choice. Here, I turn from the human component of the network to the addition of the computer.

## **2.5 The computer as disruptive technology**

As the process of collecting information and converting that information to knowledge can, of course, take place offline, does the addition of computers enable better information to emerge from the networked crowd? If so, what makes the computer the disruptive technology within the network?

Kleine (2010) has shown how ICT can enable easier communication with personal and professional contacts, provide access to information and knowledge, and save the technology user time. Internet discussion forums – like many online platforms – provide several resources to the user, including information, social resources (access to people users may not otherwise be

able to interact with) and educational resources. Within Kleine’s Choice Framework, these help to empower individuals to achieve their desired outcomes (e.g. in the case of my study ‘improved health understanding’).

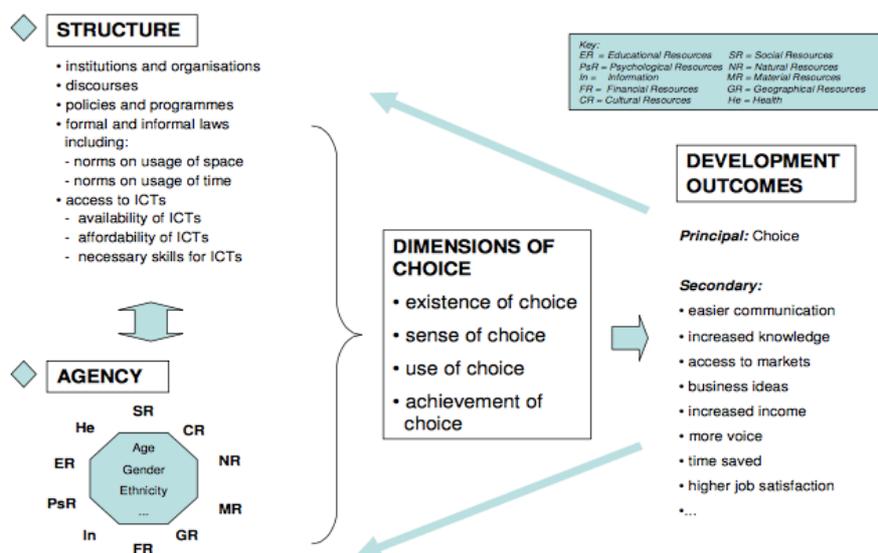


FIG 3: *The Choice Framework (Kleine, 2010, based on Alsop and Heinsohn, 2005 and DFID, 1999)*

Kleine’s Choice Framework shows how ICTs can influence, amongst others, geographical resources (enabling access to facilities, or to the experts in those facilities, even if they are far away), cultural resources (by providing a space where knowledge can be pooled or shared), educational resources (by improving access to information and experience about the health condition) and psychological resources (an increased sense of possibilities). Within this, she recognizes the value of the group within the social resource, using a definition from Bourdieu (1986) of a group as “a durable network [...] which provides each of its members with the backing of the collectively-owned capital”. Hsieh et al (2013) also identify social capital as a benefit of reddit.

If we accept that, in certain conditions, intelligence can emerge from the crowd to create a group that can be more intelligent than the sum of its parts, then the greatest resource that the Internet and Web 2.0 offers is the social resource. Computers can network more people together to create larger ‘crowds’ than ever before. In the next section, I will consider what this means for Collective Intelligence and reflect on what it might suggest for the future use of reddit and similar platforms.

### ***Computer Mediated Communication***

The term Collective Intelligence was first used in the field of computer science in 1978 (Hiltz and Turoff, 1978), to describe the intelligence that might emerge from a group of human operators linked together by computer mediated communication (CMC) – communication enabled by a computer network. If the answer is indeed ‘out there’, to be aggregated from the composite answers of crowd members, the value of platforms such as reddit is that they help us to network the crowd, providing a space in which its members can congregate and deposit their answers.

Hiltz and Turoff saw the potential of early conferencing systems such as the Department of Defence’s Advanced Research Projects Agency Network (ARPANET) and the New Jersey Institute of Technology’s Electronic Information Exchange System (EIES) to change the way people can communicate. Such systems increased the speed of communication but also the volume: before their introduction, there had been no way for a *group* (Hiltz and Turoff’s emphasis) of people to adequately exchange information among themselves and to reach collective decisions in real, or near-to-real time, other than to meet face-to-face and talk it out – a system that had essentially stayed the same since the invention of language. By the end of the 1970s, online community platforms were moving out of institutions and

starting to become available to the public, allowing more groups to form and interact through the online spaces they offered.

Though most online community platforms today are powered by web servers rather than the original Computerised Bulletin Board System software, the way in which they appear to work to the user – described by the inventor of the original software as “[P]eople [...] left messages saying they had some information of interest, and those who said they needed information discovered that other people using the system contacted them” (Christensen and Sues, 1978, p151) – has not significantly changed. As a platform type, discussion forums demonstrate considerable longevity. Web-based public forums such as Delphi Forums (<http://www.delphiforums.com/>), which started in 1983, and The Whole Earth ‘Lectronic Link (WELL – [www.well.com](http://www.well.com)) launched in 1985, are still active more than 30 years after their inception. Mumsnet ([www.mumsnet.com](http://www.mumsnet.com)) began in 2000 and Reddit ([www.reddit.com](http://www.reddit.com)) in 2005: both are more than a decade old. In a constantly changing online environment, this suggests that discussion forums are a valuable area for study as they are likely to survive the life of an academic project and far beyond. The insights gained from studying them, which may help improve their potential for supporting health information seeking in the future, do not appear to be in significant danger of being left obsolete by technology moving on.

Whether increased intelligence emerges from such a network and if it does, whether this happens in a fundamentally different way to how crowd wisdom may emerge in a non-networked offline environment, is a contested issue, however. While Hiltz and Turoff envisaged CMC driving a new equality of participation, in which everyone with access to a computer network can take part, this is not universally seen as a given. In the introduction to the first edition of *Networked Nation* (Hiltz and Turoff, 1978), Professor Suzanne Keller expressed concern that while computer conferencing systems offered a

way to potentially get beyond the “Spatial considerations [that] keep the size of interacting groups relatively small or promote a hierarchical structure in which a few participate while the many watch from the sidelines”, she suspected that the “resilience of the stratification” of society would ensure that the elite would not let go of the status quo lightly.

Nonetheless, throughout the 1990s, the term Collective Intelligence (CI) became widely used within computer science to describe the potential benefits of electronically mediated human collaboration and computer-supported cooperative work (CSCW) (Smith, 1994; Lévy, 1995, 1997). An interdisciplinary field began to emerge that drew not only from computer science (Schmidt and Bannon, 1992) but also from anthropology (Omicini et al, 2008), cognitive science (Halverson, 2002), behavioural psychology (Fisher and McKechnie, 2005), and organisational theory and management (Malone and Crowston, 1990). I will keep this breadth of disciplines in mind as I reflect on reddit’s technology and the human behaviour it supports.

## **2.6 Pierre Lévy’s Theory of Collective Intelligence**

Hiltz and Turoff (1978) showed that computer systems can overcome challenges of time and distance, increasing the speed and volume of communication, and freeing people from the need to meet face-to-face to exchange information in real time. Cultural philosopher Pierre Lévy has built on this to theorise that the way in which computers enable information to flow between individuals is not just about reducing the distance between them. It is more fundamental, to do with the increased opportunities CMC provides for more people – anyone who has access to an Internet connection – to *contribute* knowledge, which has the potential to affect society in profound ways. In 1995, he set out his Theory of Collective Intelligence in *Pour l’intelligence collective* (Lévy, 1995). Translated into English (Lévy, 1997), this reads:

*“No one knows everything, everyone knows something,  
all knowledge resides in humanity” – Pierre Lévy (1997, p14)*

Lévy, a cultural philosopher with an interest in new forms of media, saw potential for CMC to change human society, to be a truly disruptive technology (Bower & Christensen, 1995) on the scale of the development of language, writing and the printing press due to the way it can enable the sharing and transfer of knowledge. At first, his ideas might appear to be little different to those Galton, Ward and Surowiecki have put forward, or that was the basis of the Delphi Method: that members of a crowd have different pieces of information to contribute to a shared pot of knowledge. For Collective Intelligence to emerge, however, Lévy proposed that new communications technologies need to “filter and navigate knowledge, and enable us to think collectively rather than simply haul masses of information around with us” (Lévy, 1997, p10) – which draws a distinction between raw *information* and the *intelligence* that is processed from it. For reddit and similar platforms to enable CI, they must not only collect information in one place but also provide a mechanism for ordering and processing individual contributions, signpost users to the best or most appropriate for their needs, and help them find the missing pieces of the jigsaw they are trying to complete.

Online discussion forums – and reddit in particular – may have particularly appropriate technology affordances to enable health information seekers to ‘filter and navigate’ in this way. Such platforms have become vast repositories of knowledge themselves (Weninger et al, 2013) containing many millions of individual units of information in the form of the posts that have been made in their forums. More importantly, they also enable access to knowledgeable individuals who may hold required information but have not yet deposited it in the group pool (Kassing et al, 2015).

Discussion forums are particularly interesting in the context of Lévy's theory as the fundamental change he saw at the heart of the emerging digital technologies was a widening of the ability to *contribute* to society's collective store of knowledge, as well as to take from it. Prior to the emergence of the Internet, knowledge could only be 'shared' through professional publishing and (later) broadcasting companies, controlled by and open to a small elite, who communicated it *to* the public. Once the Internet began to penetrate society at large, the potential for the average person to deposit knowledge into the collective pool increased enormously. While access to such systems is by no means universal (DiMaggio and Hargittai, 2001; DiMaggio et al, 2004) and is affected by socio-economic status (Howard et al, 2010), race (Hoffman and Novack, 1998), disability (Dobransky and Hargittai, 2006), geography (Fuchs and Horack, 2008) and other factors, the barriers to publishing on the Internet are nonetheless considerably lower than they are in print or broadcast media.

New communication systems provide members of a community with the means to coordinate their interactions. Digital technologies have the potential to bring about broader participation in decision-making, new models of citizenship and community, and reciprocal exchanges of information. The opposite of Collective Intelligence in this context is a reliance on a single agent – for example on one knowledgeable expert (Aitamurto, 2016) – rather than the “universally distributed intelligence” (Lévy 1997, p13) of many thousands, if not millions of contributors. Lévy saw this as a fourth profound stage in the evolution of human communication, the first three being the development of language, which enabled mankind to pass knowledge along a chain of individuals; writing and mathematics, which enabled knowledge to be recorded, transferred and stored independently through notational patterns, so that direct human interaction was no longer necessary to its transfer; and printing and broadcast, which enabled more members of society to receive and draw from that recorded knowledge. The key change in the digital age, which

Lévy calls Knowledge Space, is that far more people can now give knowledge to the collective pot as well as draw from it.

Lévy's work makes an important distinction between shared knowledge, which is information *known* by all members of a community, and Collective Intelligence, which is knowledge *available* to all members of a community that can be retrieved when needed (Jenkins 2002). Collective Intelligence relies not only on the information itself, but on the existence of spaces in which it can be stored, shared, reflected on, discussed, debated and reshaped, and requested. Early CMC systems such as EIES, ARPANET, EMISARI and modern day forums such as reddit and Mumsnet, may offer the kind of Knowledge Spaces CI needs to emerge and thrive. As I explore these spaces and consider how their configuration enables and influences the process of depositing information, storing information and enabling the crowd to retrieve it or someone who holds it when needed, I will ask how such spaces might be better configured for use in future contexts.

### ***Realising the potential of Collective Intelligence***

Recognising the potential benefits of harnessing CI does not automatically mean this potential will be realised, however. As with the concerns immediately raised over the more collaborative and equal future envisaged by Hiltz and Turoff (Keller, in Hiltz and Turoff, 1978), criticisms of Lévy often point to his vision being too idealistic. The design, structure and availability of Web 2.0 technologies is likely to be at the mercy of the corporate companies that own them (Lovink and Rossiter, 2009) and, like Keller argued, it may be naïve to assume that the current power holders will not act to protect their interests (Jenkins, 2002). Web 2.0 can be an independent source of power and critique (Green and Jenkins, 2009, p215), the control of which political and

media elites will be loath to let go. I will reflect on this when I consider who owns, operates and controls reddit, in Chapters 7 and 8.

Lovink and Rossiter (2009, p10) argue that you cannot turn “tired and boring individuals into cool members of a mythological Collective Intelligence [...] if you’re not an interesting individual, then your participation is not really interesting” and Jenna Pack Sheffield (2013) also argues against Lévy’s implication that everyone has valuable knowledge to contribute, as he does not take into account the different ways knowledge is formed, protected and discussed in various disciplines and cultures. Early explorations of online knowledge exchange by MacDonald (1998) concluded that the utopian aspirations of online communities were constantly being challenged by unequal experiences, levels of expertise, access to community resources, access to community institutions and degrees of investment in traditions and norms. While Lévy envisages a Knowledge Space community developing its own set of ethical standards and articulating mutual goals, studies of online science fiction fans (Jenkins, 2002) – the most common type of early online community – found that even relatively small communities often splintered into subcommunities with narrower interests, depending on what knowledge is more valued by that community. As reddit has evolved from a single platform to a series of topic-specific subreddits, the power structures within this would be interesting to explore, though the resources to do so were sadly outside the scope of this project. In Chapter 6 I will, however, consider how the ordering of information into subreddits helps users to filter and navigate the knowledge they contain.

The emergence of Collective Intelligence appears to benefit from a certain degree of diversity within the group (Woolley et al, 2015), though too large a spread can cause more challenges than advantages: a Finnish government experiment to crowdsource contributions to a proposed reform to off-road

traffic policy (Aitamurto, 2016) drew contributions from a wider audience than would otherwise have contributed, and elicited a much wider diversity of opinions, but the policymakers found it difficult and impractically time consuming to absorb and process these often opposing viewpoints, ultimately leading to the project's failure. As this thesis unfolds, it will be important to reflect not only on the potential advantages of platforms that enable CI, but also the challenges they raise, and if it is possible to design-in mitigation strategies to overcome these.

### ***Processing Collective Intelligence***

One of the key challenges to fully harnessing the potential of Collective Intelligence will be learning how to process the information the crowd provides. The Internet and social media may provide access to collaboratively contributed and sometimes dynamic knowledge, but most is ultimately quantitative (Álvaro, 2014) – unrelated pieces of information that have been collected together in one place, such as videos on YouTube or links to news items on reddit. The key challenge for the 21<sup>st</sup> century is to understand how this can be (better) processed into qualitative intelligence by the human components of the network. Lévy believes that we have not yet developed effective (cyber)spaces in which we can form concrete realisations of this aspect of Collective Intelligence, for example by creating a decision-making process in which every citizen can take an active part. We have therefore not (yet) reached the potential Collective Intelligence offers.

It is perhaps better to see Lévy's theories as offering an ethical yardstick for contemporary developments (Jenkins, 2002), which frame a vision of what the Internet *could* offer, rather than what it does. While his approach has been used to explore how Collective Intelligence emerges from groups (Woolley et al, 2015); how networks can be organised to maximise their CI potential

(Malone et al, 2008; Broadbent and Gallotti, 2015); in education, to achieve open peer review (Cohen 2010; Fitzpatrick, 2011; Sheffield 2013); in organisational decision making processes (Introne et al, 2013; JafariNaimi and Meyers, 2015); in online gaming (Morreale and Bertone, 2015; JafariNaimi and Meyers, 2015) and in policy formation (Introne et al, 2013, Aitamurto 2016), Lévy himself would argue we have not yet developed the appropriate notation and organisational systems to be able to abstract the information, nor retrieve it, in the most efficient way.

Nonetheless, Jenkins (2002) argues that Collective Intelligence expands a community's productive capacity by freeing individual members from the limitations of their memory and allowing the group to act upon a broader range of expertise. Lévy believes that for the first time in history, humanity is growing a universally interconnected common memory, for which the key skill becomes organisation and management, rather than retention of knowledge. This organisation and management structure may be Surowiecki's 'clever mechanism' by which we will be best able to harness the wisdom of the (networked) crowd, but what it should look like is a gap in our knowledge.

I ask whether Internet discussion forums can provide that 'clever mechanism' and if so, how. They appear to be well-placed to do so: they are part of the World Wide Web that connects every human being with access to a computer to every other human being with access to a computer – in 2015, this stood at 3.2 billion people, approximately half of the world's population (ITC, 2015) and in the UK, 45.9 million (87.9% of the adult population) use the Internet regularly (ONS, 2016). The inequalities of access discussed above notwithstanding, all these individuals are free to make contributions to the collective pot of human knowledge.

Discussion forums such as reddit have their own internal organisation and management – consisting not only of the site owners and administrators, the forum moderators and the users, but also of the norms and conventions that develop amongst the forum users. They have an internal ordering system: the arrangement of discussions into specific topics, presented as subreddits and ‘threads’, that help with navigation and retrieval, and the cost of entry is reasonably low. They are accessible to a vast crowd of potential users.

### ***How Does Collective Intelligence Emerge?***

Collective Intelligence is clearly not just about the numbers of people involved however: the outcome of a joint activity needs to be better, not just bigger, than a project comprised of individual contributions. Projects that have attempted to apply theories of CI do appear to turn the sum of individual intelligences into a qualitatively distinct phenomenon. For example, CI has been applied to create maps of remote areas of the world, by networking online mappers working remotely from satellite images with local mappers on the ground, achieving more accurate maps, in a much shorter time, than has been achieved by any other means (Broadbent and Gallotti, 2015). An open peer review process by CI trialed by the academic journal *Shakespeare Quarterly* in 2010, resulted in all seven papers that were put out to review being eventually accepted for publication (Cohen, 2010). Climate CoLab (Malone, 2011) crowdsources solutions to climate change challenges and has generated proposals of substantial novelty as it has evolved (Introne et al, 2013; Duhaime et al, 2014). These studies suggest that the addition of computers to the human network has a profound and unique effect.

### ***Notation and Artefacts in Knowledge Exchange***

To explain what this effect might be, I turn to literature from the field of cultural theory (on which Lévy drew) and consider how language, writing and

complex notation systems have driven cultural and socio-economic development. I will first introduce this, and then show how computer networks and online discussion forums fit into this framework.

In *The domestication of the savage mind* (1977), the anthropologist Jack Goody describes how the development of writing allowed humankind to abstract our thoughts, creating the basis for new forms of classificatory thinking which ultimately led some societies to develop systems of administration, trade and agriculture. Maps, charts, graphs and tables have enabled us to abstract and compare the things they represent in a way that is not possible without such representations (Latour, 1986). Once ideas can be represented, the representations can be reflected on, leading in turn to new ideas. Being able to ‘materialise’ ideas by writing them down and storing them in some way makes them more stable (as they are less likely to be lost or corrupted in retelling and translation) and enables them to be transmitted not only from person to person, but also across time and space, requiring networks to include knowledge depositories such as libraries or the Internet, in which knowledge is stored and from which transmitted knowledge can be retrieved. The Internet provides transmission and storage components, but we also need to make sure this information is contemplated and reflected on, rather than just collected and filed away (Carr, 2011).

In discussion forums, not only the information itself, but often the discussion, debate and thought processes that developed it are recorded for posterity. Such forums often have unique forms of notification that have developed amongst their users, such as upvoting and downvoting, or signposting that signals relevance to specific topics or indicates specific qualities of the posters. The voting system on reddit (covered in detail in Chapter 4) enables users not only to be able to see which answer the community considered the best, but in some cases, how and why it arrived at this decision. Can online discussion forums therefore be considered a new form of notation and artefact, that enables us to

represent our knowledge, reflect on it, and develop it in ways that were not previously available? This has been reflected on in respect to the Internet in general (Crystal, 2001; Baron, 2010) but not with regard to discussion forums or reddit in particular, as I aim to do.

CI may emerge most effectively when the computer network is considered as an integrated part of the human mind, which extends and enhances that mind but is not separate from it. Philosophers Clark and Chambers (1998) argue that the human mind should not be thought of as being contained only within the skull: it can be actively coupled with external objects, such as pen and paper or the nautical slide rule, to enable it to perform more complex tasks. These externalities should therefore be considered part of the cognitive process.

A key notion in CI is that collective knowledge needs to be readily available for retrieval. There is too much knowledge in the world for one human mind to retain and so we need to work collaboratively if we are to realise the potential it offers. Broadbent and Gallotti (2015) argue that to understand Collective Intelligence, we should look at how human collaboration has evolved offline. Early humans began to pool their mental attitudes and skills once they realised they could not respond to increasing environmental pressure alone (Tomasello, 2009a, 2009b and 2014), overcoming individual limitations to construct artefacts that enabled knowledge to be shared across people, time and space. The Internet may be the next step on this developmental path.

Edwin Hutchins, one the main developers of the Theory of Distributed Cognition, sees significant parts of the computational work involved in any activity as being carried out in interaction with external systems, indicating that cognitive processes are distributed amongst people, artefacts and processes (Hutchins, 1995 and 2006). Artefacts and processes are an integral

part of bringing about Collective Intelligence, creating a collective or shared *intentionality*, “a uniquely human sense of ‘we’” (Tomasello, 2009a, p57) that leads to Collective Intelligence. Broadbent and Gallotti (2015) suggest that while correctly configured Internet spaces can enable this shared intentionality to evolve in geographically distanced individuals, it is dependent on the end goal required of the group being clear. If it is, the tools with which the online collectives interact and work become part of the cognitive creative process, extending the mind both individually and collectively (Clark 2002). I will explore how, and in what ways, reddit has approached and might achieve this abstraction and retrieval in Chapters 4 and 6.

If one accepts that the computerised Knowledge Space is the transformative element that provides the potential for Collective Intelligence to emerge, the next question to consider is, how is that space best configured to enable the human elements of the network to come to, and act in it, most effectively? How do online forums identify their most likely optimum configuration?

## **2.7 The Collective Intelligence Genome**

Surowiecki believed that the Wisdom of Crowds depends on each contributor being able to act independently, so that they do not fall foul of groupthink and cognitive bias. To be able to answer more than just quantifiable questions – to move towards being able to engage in more deductive and critical reasoning, which critics of the Wisdom of Crowds consider is not currently enabled (Lanier, 2010; Schuurman et al, 2012) – the members of the group need to be able to communicate and collaborate, but ideally in a way that does not bias them. Collective Intelligence depends on more than providing a similar (cyber)space in which different people can act independently (Malone and Bernstein 2015, p3). How this space is shaped depends on what exactly the group that meets in it is trying to achieve. This can be categorised using the

Collective Intelligence Genome (CIG) (Malone, Laubacher and Dellacros, 2009). Within this model, different characteristics of platforms and systems are analogous to genes, able to be recombined in different ways to produce the combination, or genome, most suited to the purpose for which the application is intended. The CIG has four basic genes: What?, Why?, Who? and How?

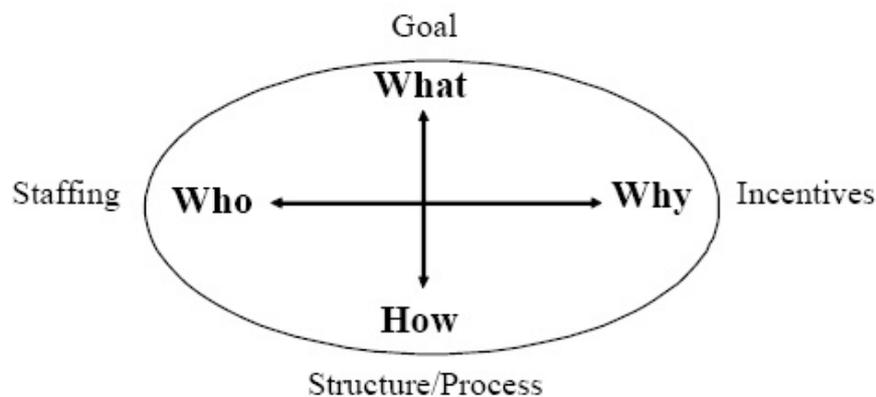


FIG 4: *The Collective Intelligence Genome (Malone, Laubacher and Dellacros, 2009)*

While terms such as Collective Intelligence ‘Genome’, and its reliance on only four basic concepts might be considered overly simplistic, it does help one to visualise how four basic building blocks – like the A (adenine), C (cytosine), G (guanine) and T (thymine) nucleobases that make up DNA – can be recombined in different ways to present very different results. Another strength of the analogy is that these ‘genes’ – like real genes – have different variants, called alleles in genetics. Human genes for eye colour have alleles that combine in different ways to result in blue, green or brown eyes, for example. The CI ‘genes’ also have variants: e.g. the CI ‘gene’ for Why? has three: Money, Love and Glory, by which the motivations of the owners, operators and/or users of the platform can be (very broadly) categorized and, as with biological genetics, there can be a spectrum between the variants rather than absolutes.

The motivation of Why? may be a combination of love, glory and money, and different users may have different motivations, but there is some advantage in knowing which is the intended motivation for platform users, particularly as this will influence the PEOU and PU of the system. By understanding which CI genes are useful in which situations, it may be possible for systems to be ‘genetically engineered’ to be as effective as possible for specific tasks. This approach might also be useful for identifying which existing systems might be most appropriate for any given context.

COLLECTIVE INTELLIGENCE GENE	VARIANTS
What (is it intended to be)?	<i>Collection</i> (e.g. YouTube)
	<i>Collaboration</i> (e.g. single Wikipedia entry)
Who (runs it)?	<i>Crowd</i> (equality of participation)
	<i>Hierarchy</i> (some contributors have privileges over others)
Why (do they do it)?	<i>Love</i> (altruistic)
	<i>Money</i> (reward)
	<i>Glory</i> (power)
How (by what process)?	<i>Create</i> (by <i>Collection</i> , contested or uncontested, or <i>Collaboration</i> )
	<i>Decide</i> ( <i>Group</i> decision by <i>Voting</i> , <i>Averaging</i> , <i>Consensus</i> or <i>Prediction Market</i> ; or <i>Individual</i> decision by <i>Market</i> or <i>Social Network</i> )

TABLE 2: *The Collective Intelligence Genome and its Variants (after Malone, Laubacher & Dellacros, 2009)*

The CIG model is far from perfect: it is obviously simplistic, and does not propose how equality of participation might be achieved, for example. In the same way as Lévy’s theory is perhaps best taken as a framework that suggests the potential of Collective Intelligence, the CIG should be taken as a starting point to help developers of a platform consider what they want that platform to achieve. Should it be a space to *Collect* things in one place or to enable

*Collaboration?* Is the aspiration for all users participate equally (even if in practice this may be not be realised) or are additional privileges for some users built into the system? In which parts of the system is it (more) important for users to be (more) equal? Harking back to User-Centered Design (Norman, 1986), what psychological goal does it aim for? Does it need to be fun to use to ensure its perceived ease of use is greater than its perceived usefulness?

The CIG of Reddit has not yet been mapped, but this is something I intended to do as part of my study. In Chapter 5, I identify some of the likely user requirements of a platform intended for use during a public health emergency, and I will orient these to technology affordances of reddit to assess its likely 'genetic fit' as a Knowledge Space for exchanging health information during such an event. The CIG acknowledges that there are different forms of Collective Intelligence and different ways for it to emerge: the most appropriate genome may differ depending on the context and end result(s) required. In the case of health information, can understanding the best CIG help give rise to online discussion forums that are not just collections of information but collaborative spaces from which Collective Intelligence can emerge? Does the system offer the flexibility that might be needed to achieve different results for different users in changing and fluid contexts? Does this help or hinder such spaces to become the 'clever mechanism' through which Internet users might be organised and managed, enabling the efficient retrieval of the knowledge the online space contains or provides access to? Theories of Crowd Wisdom and Collective Intelligence would suggest that the crowds of online discussion forums may collectively provide access to 'better' health information than could be obtained from a single source. They may enable users with limited knowledge to pool what they have and thus improve the overall expertise of the group. Coulson and Shaw (2013), in a study of online support groups for cancer patients, described such groups as a 'communal brain' that provides the potential for members to access a wealth of factual and

experiential information, advice and support. Is this also likely to be true for a forum designed for use, or used, during a public health emergency?

## **2.8 Conclusions**

I offer this theoretical framework as basis for my research design. Concepts of crowd wisdom proposed by Galton and Surowiecki's ideas on the Wisdom of Crowds are developed through Lévy's Theory of Collective Intelligence to inform our understanding of what technology affordances computer mediated communication offers and what potential this holds (Baccarne, 2012; Schuurman et al, 2012). I draw on ICT for Development (ICT4D) literature to consider the resources ICT offers (Kleine, 2010) to improve the social capital of individual users through access to a wider user community, and to consider whether, when the ideal solution is not available, something worth having might still be achieved (Chambers, 1981, 1992; Sen, 1999). These concepts have not, so far, been related directly to online discussion forums, particularly regarding the sharing of health information as I propose to do in this study. Surowiecki does offer the pooling of information by laboratories during the SARS outbreak as an example of the Wisdom of Crowds, but this information exchange was not carried out in public online discussion forums.

A greater understanding of how the affordances of online health discussion forums are influenced by its system architecture and the psychological goals of the user may help us to better determine what source, message and channel characteristics add (perceived) value, improve PU and PEOU, and straddle the execution and evaluation gulfs, in turn determining the extent to which the technology is not just valuable but also likely to be valued by the user. This will also help to improve design, and identify what existing spaces might be best suited to this kind of health information exchange. Cultural theories on the development of human collaboration and systems of notation also add

value to how this might be understood, suggesting ways in which the nature and processes of collaboration need to be considered in tandem, so that systems do not simply become dumping grounds for information that cannot easily be retrieved and used when needed.

Bonabeau (2009) sees the Collective Intelligence of the Internet as enabling two key functions: [1] Generation of potential solutions and [2] Evaluation of potential solutions – essentially the *Create* and *Decide* of the CIG's How? The Collective Intelligence enabled by the Internet might help us to not only make the fast decisions our ancestors required for survival (Tomasello, 2009a, 2009b) but also to quickly explore multiple possibilities, consider the potential opportunities and challenges of each (Carr, 2011), and to come up with accurate responses within a short timeframe – characteristics that may be particularly relevant during a health emergency.

I therefore suggest that the Theory of Collective Intelligence and the CI Genome provide a helpful framework for understanding how computer mediated communication might be configured to enable decision-making, as well as just information sharing. This includes whether and how such information sharing might differ under different conditions such as 'Business as Usual', or a 'State of Exception' (Agamben, 2005) when the usual norms and laws of society are suspended or become flexible. I would argue that a State of Exception was observed during the Ebola PHEIC, particularly in the Ebola Treatment Centres and in the areas of Sierra Leone's Freetown that were placed under military-enforced quarantine at some stages of the outbreak. Can discussion forums provide spaces where options can be discussed, and potential solutions presented and debated? Can comments and the replies to them evaluate those solutions, further aided by the voting mechanisms that enable the online community to collectively decide what action should be taken?

I aim to contribute to existing knowledge by exploring the CI Genomes of existing Internet discussion forums and considering how this understanding may help to identify those of most benefit to health discussions. How might such forums generate potential solutions to health challenges and enable those potential solutions to be evaluated? And what if any modification(s) may help or hinder their effectiveness in future health emergencies?

To my knowledge, my study is the first to consider how Lévy's Theory of Collective Intelligence and the MIT-developed Collective Intelligence Genome, apply to online health discussion forums and to reddit specifically.



### 3: METHODOLOGY

---

#### 3.1 Introduction

The methodological approach I have taken throughout my study could not have arisen from one discipline alone. I have been funded by a Royal Holloway Reid Scholarship in Health, the Human Body and Behaviour (H2B2), administered by the School of Biological Sciences and co-supervised through the Department of Computer Science and the Department of Geography. All three disciplines have informed the research design described in this chapter and the research methods chosen for it. My own research background has further influenced my approach. From a first degree in Biological Anthropology, through nearly 20 years as a journalist and 10 as a policy researcher, I have undertaken projects investigating biological threats on the UK's National Risk Register (Cole, 2013 and 2016) and barriers to the uptake of emergency services IT systems (Cole, 2010; Cole and Hawker, 2014). This has informed my preferences and approach to design study: not least, it has highlighted the value of mixed methods, of moving quickly before opportunities are lost and of talking directly to people with personal experience of the topic being studied – in this case, of living through an outbreak of serious disease and of discussing that outbreak on Internet forums.

I consider my approach to be interdisciplinary, “involv[ing] the use of an innovative conceptual framework to synthesise and modify two or more disciplinary approaches to deal with a research problem” (Graybill et al, 2006), distinct from cross-disciplinary, in which “researchers from two or more disciplines work [...] collaboratively on a common problem without modifying disciplinary approaches” (*ibid*). Interdisciplinary approaches afford new and exciting opportunities. This can inevitably involve some compromise on the depth and perceived ‘completeness’ of the finished work when

considered by a single-discipline reader (Blackmore and Nesbitt, 2008) but this is off-set by the greater breadth of understanding gained from approaching a topic from more than one angle.

The aim of my study has been to understand how discussion forums may help or hinder health information seeking during outbreaks of serious infectious disease. To achieve this, I have drawn on methods that have been used in computer science to understand how systems can be best configured to deliver user requirements (Norman, 1986, 1988, 1999, 2002; Davis, 1986 and 1989; Ventakesh et al, 2003; Ventakesh and Bala, 2008). I also draw on methods used by evolutionary theorists to explain how and why humans work collectively (Tomasello, 2009a) and to understand the role played by artefacts in the development of human cognition, communication and cooperation (Goody, 1977; Gibson, 1979, 2014; Latour, 1986, 1991; Seifert and Hutchins, 1992; Lévy, 1995, 1997, 2010; Clark and Chambers, 1998; Clark, 2002). I draw on their approaches to understand not only why and how discussion forum users are able to use the technology available to them, both individually and collectively as members of online communities, but also why they might want to. This led me to design a programme of research that not only examined how the technology works, but also why the human operators are motivated to use it, and what challenges to utilisation they may face. This required a deeper examination of social science methods and literature than would be required for a project rooted purely in Computer Science.

Correspondingly, my data collection has taken a mixed methods approach. Qualitative data was gathered from semi-structured interviews, free text boxes on surveys, and observation of conversations taking place on public and private discussion forums. Quantitative data was collected from surveys, reddit traffic statistics and logs recording moderator actions. Mixed methods is considered to provide a more complete understanding in situations where

quantitative data may not give the full picture (Tashakkori and Teddle, 1998; Creswell, 2013, 2014). I felt this was particularly appropriate to my study as I not only wanted to know which information doctors rated good or bad, but also what they considered to be good or bad about it. I wanted more than just the number of moderator actions undertaken on a health forum on a certain day; I wanted to know how the moderator felt about their role and whether they considered any of the actions to be more important or valuable than others. Mixed methods has also been identified as particularly useful for tying together several steps in an evaluation process (Creswell, 2014), as qualitative approaches can be used to probe figures and statistics to understand why one value may be higher than another, or to try to explain unexpected or apparently contradictory figures, for example by asking respondents why they chose a certain option on a multiple choice questionnaire, and enabling the researcher to reorient later research if necessary.

The discussion forums I studied present a permanent record of activity during specific events, including the numbers of posts made per day and the rate at which the forums gained subscribers. Mixed methods enable me to triangulate this quantitative data with qualitative data gained from interviews conducted with forum users to see, for example, if their memories of which event(s) triggered an increase in activity on a forum is upheld by increased numbers of posts corresponding to the date of the remembered event, in which reported discussions are correctly recalled. Reported behaviour is not always reliable (Alshenqeeti, 2014), but using mixed methods provided a valuable opportunity to triangulate reported behaviour against the permanent records provided by the forum, moderation logs and traffic statistics.

The qualitative and quantitative data I collected was gathered across four separate studies, each detailed below. Analysis is spread across four results chapters (Chapters 5-8) but as each chapter covers themes that emerged across

all four as they progressed, rather than each presenting the results from a single study, I will cover the methodology used for each study in detail here. Sufficient detail is given to enable my research to be reproduced.

My study design was considerably influenced by the outbreak of Ebola in West Africa that started in early 2014, after I had commenced my programme of research. While it led to an horrific loss of life, the Ebola outbreak also provided opportunities to study an emerging outbreak in real-time and to connect with people who were genuinely affected by it. Without this, my research would have been based on theoretical scenarios and retrospective records. The precise methodologies I follow are to some extent experimental and emergent, but this has been identified as particularly appropriate for situations that are, “unchartered, contingent or dynamic” (Charmaz, 2008, p155). In responding to this opportunity, I have also kept in mind Robert Chambers’ approach that when time is constrained it is better “not to wait, but to start, stumble, self-correct and then share” (Chambers, 1992, p56).

My sample sizes are relatively small, and not especially conducive to statistical analysis, which I recognise as a limitation of the study, but the quality of the respondents providing the data mitigates this. Working doctors have very little free time, but 35 contributed to the data in this study across two separate surveys. I drew from a small pool of forum moderators who have unique experiences: not including myself, just 13 people have been moderators of r/ebola: I interviewed five and corresponded with a sixth.

Four separate studies across three distinct phases combine to provide the data I will present and analyse over the coming chapters.

- **Phase I** consisted of interviews with Ebola witnesses – NGO and international workers based in areas of West Africa affected by Ebola.

- **Phase II** consisted of two survey questionnaires to doctors on the quality of online health information
  - Phase II-A (the pilot study) asked respondents to rate health information across three online platforms
  - Phase II-B asked respondents to rate information on r/ebola
- **Phase III** consisted of interviews with reddit moderators, plus observation of activity on reddit, including data from areas of the site not available to general users.

### 3.2 Phase I: Interviews with Ebola witnesses

The aim of the Phase I study was to gain a sense of what questions people had during a PHEIC, where they looked for answers, and whether their health seeking behaviour changed during such an event. This included whether they looked for new sources of information or used new channels to access it.

Between 8<sup>th</sup> July and 5<sup>th</sup> November 2014, I interviewed 14 employees of international non-governmental organisations (NGOs) and international businesses with operations in West Africa. Ten of the interviewees were (or had very recently been) stationed in Liberia and four were in Sierra Leone; Eleven of the interviewees were women, three were men and they ranged in age from mid-twenties to late-fifties. Six were British, seven from the U.S. and one was Dutch.

The cohort was recruited mainly through a request posted on the Liberia Expats GoogleGroup, a members-only online forum for people (mainly British and American) who are, or who have previously been, working in Liberia. A brief description of the aims of the research was posted on the message forum, to which members who wanted to volunteer themselves could respond.

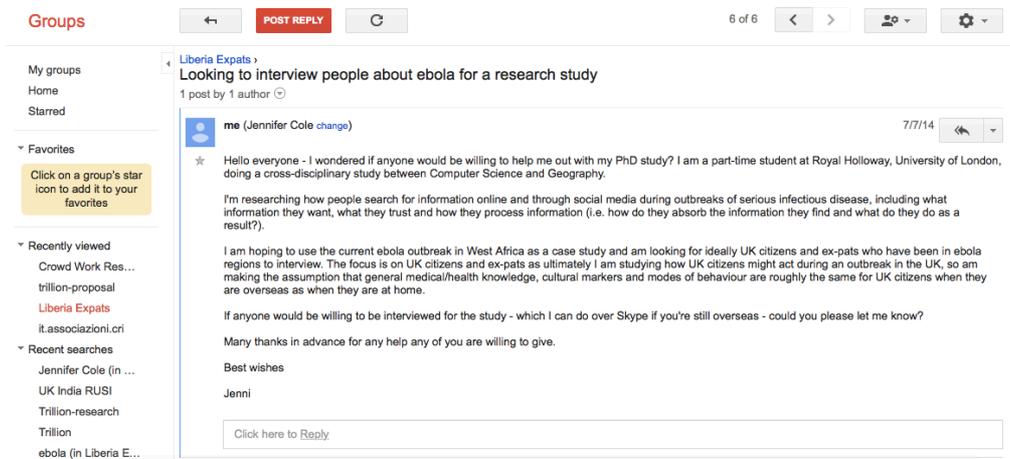


FIG 5: A screenshot of the message posted on the Liberia Expats GoogleGroup on 7<sup>th</sup> July 2014, asking people to volunteer to be interviewed.

The participants from Sierra Leone were recruited through approaching people known by me or my colleagues to be currently working in the affected region(s), a technique known as purposive convenience sampling that targets specific individuals known to have the qualities required and who are easily accessible to the researcher, rather than aiming for a random sample of the available population (Etikan et al, 2016).

Royal Holloway's Ethics and Risk Assessment requirements precluded me travelling to West Africa to conduct interviews in person, as the threat from Ebola was considered too great. I was also discouraged from approaching staff working for medical NGOs, in case asking them to participate took time away from life-saving activities. The interviewees, therefore, came from a combination of conservation NGOs, Christian NGOs offering education and health support to disadvantaged communities, and from the administrative staff of larger NGOs such as UNICEF, which were thought to have the logistical capacity to participate in the study without impacting their essential frontline services.

Study inclusion criteria were that the interviewee had to speak English, so that they could be interviewed verbally, and either be in one of the Ebola-affected countries (Liberia, Sierra Leone or Guinea) or have returned from one of them since the outbreak began. I acknowledged that this selection criteria, and the recruitment methods undertaken, is likely to have recruited individuals who represent an international elite that is somewhat distanced from the typical population of the countries in which they resided. NGO workers are often more socially elite than the population they serve or the population they come from (Mitra and Van Delinder, 2007; King 2015). They tend to be highly educated and middle-class; their health literacy and research skills are likely to be above average and this may have influenced their responses. It is equally important not to assume that because most of them were U.S. and U.K. citizens, their responses provide an accurate representation of the responses likely to be given by such people during an outbreak of serious infectious disease on their home soil. Nonetheless, the answers they gave, which related to their own information seeking behaviour as well as that of acquaintances, colleagues and friends, provides some indication of the concerns that may arise during such events.

In total, 14 interviews were undertaken, conducted over Skype with two exceptions: one was conducted face-to-face with an interviewee who had recently returned from Liberia to her home in Washington DC and was available for interview when I was also in the city; the other was completed by e-mail after the Internet connection proved too poor to sustain a Skype session. Interviews were audio-recorded (using iFree Skype Recorder software) and transcribed by me (except the interview conducted by e-mail). The transcriptions were sent to the interviewees for verification and amendment. Interviewees were given the option at this stage of adding in additional points not covered during the interview itself. The recorded

interviews vary in length from 13.5 minutes to 69.5 minutes, with an average length of 35 minutes.

The identity of the interviewees is protected by coding the participants as WAG (indicating West Africa Group) and with the numbers 01-14, indicating the order in which the interviews took place.

While a larger sample would have been preferable, resources available for the study, ethical considerations and the timescale available imposed restrictions. I acknowledge this may limit the robustness of the study and may bring some of its conclusions into question. Further research during a future disease outbreak will be needed to verify and test these findings.

Findings from this, Phase I of the research, are presented mainly in Chapter 5, Information Requirements During a PHEIC. The findings were also written up, submitted to and published in the *Journal of Business Continuity and Emergency Planning*, Vol 9., No 2, Winter 2015-2016, ISSN 1749-9224.

### **3.3 Phase II – Quality of information in online health forums: Semi-structured questionnaires sent to health professionals and patients**

The second phase of the study aimed to determine whether the quality of information found in online discussion forums is sufficiently high to consider them worthy platforms through which to encourage public health messaging and health information seeking during a public health emergency. Previous research into what constitutes ‘good’ information in online discussion forums has tended to focus on content or content contributors who are well-received by the community (Ma et al, 2012; Das and Lavoie, 2014; Kassing et al 2015). These studies have not assessed the quality of the information beyond using popularity (decided by a post’s vote score, for example) as an indication of ‘quality’ (Lampe and Resnick, 2004; Stoddard, 2015a, 2015b) and make no

attempt to determine whether the popular information is accurate, while others (Impicciatore et al, 1997; Whitelaw et al, 2014) have matched information found online against professional medical publications for accuracy but have not determined if the more accurate information is more popular. This inspired me to survey doctors directly for this phase of the study and to ask not only whether they thought information was ‘good’ or ‘bad’, but also why they made those judgements, and in what ways they thought ‘bad’ information could prove problematic for those receiving it.

Phase II of my study contributes to the existing literature by determining the quality of health information found online, based not only on whether it conforms to medical convention but also whether doctors consider it to be appropriate or harmful. Such analysis is missing from the current literature.

The study was undertaken in two parts: Phase II-A involved UK-qualified medical doctors and UK (London)-based non-medically qualified individuals assessing the information found in three online discussion forums (reddit, Mumsnet and Patient.co.uk) relating to three health conditions (diabetes, chickenpox and HIV). Phase II-B involved UK-based medical doctors attending a conference held in London being asked to rate information given in response to questions asked on the reddit Ebola discussion forum r/ebola.

### ***Phase II-A: Investigation of information quality in online health forums***

For Phase II-A, I selected three health conditions that affect a high number of individuals in the UK: diabetes, chickenpox and HIV. According to the most recent figures from Public Health England, an estimated 107,800 individuals in the UK were living with HIV in 2013 (Public Health England, 2014b). An estimated 3.2 million (7%) of the UK population is living with diabetes (Diabetes UK, 2015), of whom 10% have Type 1 diabetes and the remaining

90% Type 2. An estimated 90% of all Britons will have had chickenpox by the age of 15 (Public Health Wales, 2015), though no exact figures on infection exist for the UK as not all cases receive clinical attention.

When I commenced the research programme, there was evidence of a high volume of health seeking information taking place in relation to all three conditions. Diabetes and HIV both feature in the top 10 ‘Most Searched for Diseases’ on Google (diabetes at number 2, with more than 9 million monthly global searches in 2013; HIV at number 4 with more than 6 million, and AIDS at number 6, with five million – PharmaForward LLC, 2013). While chickenpox appears lower on the list (at 43, with over half a million global monthly searches), it is one of only a handful of communicable diseases listed and is the most significant childhood disease in the UK.

It seemed likely that a considerable volume of online health information would exist for these three conditions, and that forums where health is discussed were likely to have discussion threads related to them.

### **Selection of websites and discussion forum threads**

Three online discussion forum websites were selected based on their popularity and common usage by the UK population (rather than amongst specialist interest groups or social media super-users): two general discussion websites: reddit, which in March 2015 had 6.6 million unique UK users per month, or one in 10 of the UK population (Morse, 2015) and Mumsnet, which had a reported 7.5 million registered users in March, 2015 (Roberts, 2015) plus one health-specific site (Patient, which receives 18 million visits a month<sup>2</sup>). I investigated each of the three websites to see if their message forums had existing discussion threads related to these conditions and found that all three health conditions are indeed discussed on all three forums.

---

<sup>2</sup> According to its own website, last accessed 23<sup>rd</sup> March 2017

I selected specific discussion threads for the survey subjectively by undertaking a basic search inside each selected website on the chosen health conditions between 15<sup>th</sup> and 17<sup>th</sup> February 2015, and reading through the returned results to find questions for which I reasoned the original poster could (and probably should) have sought advice from a qualified medical practitioner. Discussion threads were rejected if the question did not require a medical or scientific reply (e.g., a diabetic asking others whether they thought disclosing his condition on job applications would disadvantage him) and if the question had received less than two replies. I selected 25 questions I considered suitable (Reddit: n=9, Mumsnet: n=8, Patient: n=8; diabetes: n=8, HIV: n=9, chickenpox: n=8) according to the order they appeared in the search results, working backwards from the highest posts, which were those most recently posted on Mumsnet and Patient, and the highest scoring on reddit. Each question and the discussion thread following it was then assessed by more than one assessor. In total, 79 assessments were returned by 17 assessors. Each question was assessed by a mean of 3.2 assessors (range 2-7).

### **Selection of study participants**

I aimed to have the information in the forums assessed for quality by UK-qualified medical doctors and, in the pilot study, also by UK individuals who were not medically qualified but who have experience of the health issue being discussed as a patient or carer. I recruited six doctors through Ashford and St Peter's Hospital, which has links with the Health, Human Body and Behaviour programme at Royal Holloway. Two other doctors, known to me personally, were also invited to participate. The group included general practitioners, hospital infection specialists and diabetes consultants.

Nine non-medically qualified participants were also recruited. Four were reached by contacting the coordinators of two (offline) support groups for diabetics, via contact details given on the website of the diabetes support

charity Diabetes UK. Five parents of children in the common age-group for contracting Chickenpox (2-10 years) were recruited through the Parents and Friends Associations (PFA) of two local West London schools (Lovelace Primary School in Chessington and Putney Girls High School). The Terrence Higgins Trust, a charity supporting people living with AIDS and HIV, was also approached, but did not reply. As such, no HIV-positive patients participated and the questions relating to HIV were answered by doctors only. Participants were self-selecting and therefore may be subject to selection bias.

Demographic data collected on the participants was minimal: whether they were medically qualified was recorded and they were asked to confirm that they were over the age of 18 before taking part, but no other details were recorded as these were not deemed necessary. Participants were given the option of taking part anonymously; 12 chose to disclose no information other than if they held a medical qualification. Five doctors but only one of the non-medical participants provided a follow-up email address.

The 17 participants were emailed a list of paired URL links for each discussion thread to be assessed (a generic version can be found at Appendix II). One URL linked to the online forum discussion, where they saw the thread *in situ* with no modifications made for the sake of the study, and the other to an online assessment form. Each discussion thread was assessed against the same five criteria, with the participant responding by rating the information from, in their opinion, highest quality to lowest quality over a range of 1-5 on:

- The medical/scientific accuracy of the information
- The medical/scientific completeness of the information
- How sensible they considered the answers provided to be
- Whether they thought someone reading the website would act appropriately based on the information provided
- How useful they felt the answers given would be to the original poster.

In total, 79 assessments were returned, an average of just over three for each of the 25 questions (range = 2-7). An additional question was asked to check that the respondents found the discussions easy to follow. Four assessments (of 79 = 5%) recorded some level of difficulty in following the discussions, all in relation to long discussions on chickenpox.

The responses assessed perceived factual quality of the answer (accuracy and completeness); gave a subjective assessment on that information (How sensible was it?); and subjective assessments of how the reader might respond (Would they act appropriately, and would they find the information useful?). I included this differentiation in the questions as, while many previous studies (Impicciatore et al, 1997; Schwartz et al 2006; Whitelaw et al, 2014) have criticised online health information for being of poor or variable quality, far less have found evidence of poor information leading to inappropriate or dangerous health decisions being made (Crocco et al, 2002; Eysenbach and Köhler, 2002; Bansil et al, 2006; Nölke et al, 2015). As even fewer studies focus on whether discussion forum readers take harmful action based on poor information they find, I was interested in exploring perceptions around this.

In each assessment, the discussion threads could be assigned one of five rating values, for which the highest (1) related to the best quality information and the lowest (5) to information considered to be inaccurate or ill-advised. Criteria for marking were consistent across each health topic and website, providing a potential overall score of between 5 (5 x 1, top rating for each criteria) and 25 (5 x 5, lowest rating for each criteria) to each discussion forum thread.

Participants were invited to participate between 12<sup>th</sup> May and 4<sup>th</sup> June 2015, and given 2-3 weeks to reply. The final survey assessments were accepted on 18<sup>th</sup> June 2015. Participants were sent on average eight discussion threads to assess (each of which required assessments of the five separate criteria) based

on their experience or medical expertise, with only one participant – a recently retired GP – offered all surveys to complete. The assessments were completed, and results collected, using the free online survey software SmartSurvey.

Findings from this phase of the research are presented in Chapter 6 and more detailed findings have been published as a standalone paper in the *Journal of Medical Internet Research*, Vol 18., No 1, doi:10.2196/jmir.5051

### ***Phase II-B: Investigation of information quality in r/ebola***

Phase II-B focussed on the subreddit r/ebola, to determine how qualified medical professionals would assign ‘upvotes’ or ‘downvotes’ to answers given in response to questions asked on the forum compared with the votes cast by reddit users. This was intended to support the suggestion that the reddit community (as a collective) will upvote medically accurate information and downvote medically inaccurate or unhelpful information. If this is supported, then doctors should rate information in a similar way to how it is voted on by the (sub)reddit community. It will also help to support the idea that reddit has some technology affordances that enable Collective Intelligence to emerge by *Consensus Voting*, one of the Collective Intelligence Genome options for platform configuration. Members of the reddit community can see one another’s votes and comments, and are therefore arguably susceptible to groupthink (Weninger et al, 2015), but the doctors who rated the same questions did so in isolation – more like the way Delphi Method participants would engage and ‘protected’ from one another in the way Surowiecki considers is important for the emergence of unbiased crowd decisions. This approach may therefore give some indication – albeit limited due to its small size, and as a pilot study within the wider body of the research – of whether voting on r/ebola may have been influenced by groupthink.

### **Selection of the answers given on r/ebola**

I selected the answers for inclusion in the study by carrying out a search within r/ebola to select only posts without links to external content, as these were more likely to be ones in which users were asking a question. I then read through all the post titles and discarded the ones which either did not have a question for a title, or for which the first line of the text did not indicate the poster was asking a question. The remaining posts were then ordered using the options (explained in detail in Chapter 4) ‘top’ and from ‘all time’. This resulted in a top post<sup>3</sup> which had received 283 points (83% upvoted), and 53 comments; down to the bottom<sup>4</sup> one, which had 0 points (43% upvoted) and 12 comments.

Next, I selected only those questions I felt it was reasonable someone may ask their doctor; this discarded, for example, questions relating to why the media was covering the outbreak in certain ways, questions about what conditions were like on the ground in Africa, or questions relating to the administration of r/ebola. As I was interested in determining how different answers to the same question were rated against one another, as well as in absolute terms, I next selected in only those questions which had received more than 10 comments, and for which voting had taken place on those comments. This left 27 eligible threads.

From these, I selected threads for which the lead comment (ordered using the ‘best’ sorting option) had a score of at least five points, indicating that it received at least five upvotes for every downvote cast, and for which at least one lower comment had received either a negative score or no indication that any votes were cast on it. This produced a set of questions with replies which

---

<sup>3</sup> [https://www.reddit.com/r/ebola/comments/2irqqc/psa\\_want\\_to\\_help\\_fight\\_ebola\\_get\\_a\\_flu\\_shot/](https://www.reddit.com/r/ebola/comments/2irqqc/psa_want_to_help_fight_ebola_get_a_flu_shot/)?, last accessed 10<sup>th</sup> March 2017

<sup>4</sup> [https://www.reddit.com/r/ebola/comments/2jaj6a/what\\_conditionscriteria\\_would\\_need\\_to\\_be\\_met/](https://www.reddit.com/r/ebola/comments/2jaj6a/what_conditionscriteria_would_need_to_be_met/)?, last accessed 10<sup>th</sup> March 2017

had received mixed reactions from the r/ebola community: some replies had been received favourably (those that were upvoted and had positive scores), some may have been ignored (those that had the default posting score of +1 only) and some had been negatively received (those which displayed a score of zero or a negative score). These answers, provided by the reddit community, could be put in front of qualified medical professionals to not only see if they would agree with ratings of 'good', 'neutral' and 'bad' but also if the doctors placed the available answers in the same order of value, from 'best' to 'worst', as the online community. This is an important consideration in light of Sen's concept of maximized choice (Sen, 1997) and Chambers' concept of 'proportionate accuracy' (Chambers, 1981), as if these are the only answers available to the online community, it is worth knowing if the community can identify the 'best' of those on offer, even if the doctors might consider that answer to be of middling quality only.

From this selection of questions, I then worked downwards from the 'top' post and eliminated questions that repeated ones that had been asked before. This left nine suitable questions, which were selected for inclusion in the study. As I had ideally wanted 12, I also included three for which no comment had received negative voting, but for which four comments had nonetheless received different voting activity. In one case, two comments displayed only the default +1 vote, and in two others one comment had received higher upvoting than the others. These also provided an opportunity to see if the doctors agreed with r/ebola that none of the answers was indeed 'bad'.

Four answers to each of the questions were selected from the question's discussion thread, including the top answer (i.e. the one that had been most heavily upvoted and was therefore considered 'best' by the r/ebola community), the bottom one (i.e. the one that had been most heavily downvoted, or had received the least number of upvotes, and was therefore

considered the least best), and two from the middle section of the thread, ensuring a selection of answers that had been well received, poorly received and neutrally received or ignored by the r/ebola community for each question. The 12 questions, and four answers given in response to each were then copied from reddit into a questionnaire format so that the doctors who were asked to rate them could not tell where the information had come from. This was to prevent any preconceived biases about discussion forums, the Internet in general or reddit influencing their judgement. Each question was presented on an A4 sheet, with the four answers displayed underneath it, along with multiple choice options to score each answer 'good', 'bad' or 'neutral', according to criteria explained in the questionnaire. The order of the answers was randomised so that positioning biases could not influence the results: the answers did not necessarily appear in the same order to the doctors as they did on reddit. Below each question was a free text box in which respondents had the option of adding additional comments to explain their response. The questionnaire as presented can be found at Appendix IV.

### **Selection of study participants**

Hard copies of the questionnaire were handed out to delegates at the Royal Society of Medicine (RSM) Military Medicine Section conference 'Non-conventional Warfare: The Medical Response' on 29<sup>th</sup> September 2016 and to attendees at a meeting of the West Kent Medico Chirurgical Society on 11<sup>th</sup> November 2016. At the RSM conference, copies were left on seats at the beginning of the conference, with attendees (all of whom were doctors) encouraged to complete them during the day and return them at the end of the event. At the West Kent meeting, this approach was taken again, but attendees were also offered stamped-addressed envelopes so that they could take questionnaires away and return them later, as it was less likely they would have time to complete the questionnaires during the two-hour meeting. In total, 27 questionnaires were returned, 17 from the RSM group, and a further

10 from the West Kent group, three on the evening and seven through the post using the SAE. Demographics for the two groups are shown below:

	RSM GROUP	WEST KENT GROUP	TOTAL
<b>Gender:</b>			
Male	10	3	13
Female	2	6	8
Not given	5	1	6
<b>Age:</b>			
18-24	2	0	2
25-34	1	0	1
35-44	4	0	4
45-54	2	1	3
55-64	2	5	7
65 and over	2	2	4
Not disclosed	5	1	6

TABLE 3: *Demographics of participants in the Phase II-B study.*

Two of the doctors in the RSM group had direct experience of Ebola patients.

The doctors awarded the answers presented a score of ‘good’, ‘bad’ or ‘neutral’. These were converted to a numerical score of good = 3, neutral = 2 and bad = 1, and averaged across each group, and across both groups combined, to determine whether the medical groups collectively considered the answers to be good, neutral or bad. In total, 27 doctors rated 12 questions, each of which had four answers, providing a possible 1,296 data points (the survey in fact returned 1,140 data points, as some respondents did not rate every answer) to compare with the votes cast by r/ebola users. Findings from this phase of the research are incorporated into Chapter 6 and the full data from the study responses is presented in Appendix V.

### **3.4 Phase III: Observation of reddit and interviews with moderators**

The third phase of the study aimed to understand how reddit is structured, how it works, what information is posted on it, how problematic information is dealt with, and how the moderators of its health forums think those forums might be used during a public health emergency.

During the period for which I was to study reddit I had two options: to adopt a non-participatory role, observing reddit for the purposes of academic study but not actively participating nor acting in a way that could affect the content or the community dynamics – or I could attempt to become a functioning member of the reddit community.

Initially, I ‘lurked’ (Golder and Donath, 2004) on the site, becoming familiar with its structure and characteristics. On average, around 90% of users who interact in any way with a website do so in a non-participatory role (Nonneke et al, 2006; Singer et al, 2014) and this has also been shown to be consistent for participation in online health forums (van Mierlo, 2014). Ratios have been shown to differ greatly across different subreddits (Weninger, 2014), but data provided by r/diabetes – the exact detail of which I was asked to keep in confidence – suggests that health forums on reddit do conform to the expected ratios. Reddit actively encourages new users to lurk before posting, to become familiar with a site’s rules, norms and standard behaviours. In October 2015, I registered with the site as a user – known as a ‘redditor’. I chose the username JenniferColeRhuk, a combination of my real name, the initials of the university (Royal Holloway) to which I am affiliated and the country (United Kingdom) in which I am based. While it is possible to choose a username that contains no identifying features (Kassing et al, 2015), I felt that it was ethical to be as transparent as possible about myself: to make it clear that any interaction I had with the site was part of an academic study, meaning that I

may have a different motive for interaction than other reddit users (the reward of a PhD, equalling the ‘Money’ or ‘Glory’ rather than the more altruistic ‘Love’ of the CIG). For this reason, I felt that anonymity was inappropriate.

I began to contact the moderators of health forums to ask if I could interview them. I first attempted to contact them by posting a message on five health-related subreddits (r/diabetes, r/cancer, r/asthma, r/multiplesclerosis and r/hivaid) and r/modhelp. I sent the same message, through reddit’s internal mail system, to the moderation teams of these subreddits plus 14 other health-related subreddits (including r/ebola and r/zika). I also sent a personal message (PM) through reddit’s internal mail system to 35 health forum moderators. An example of the message is shown in FIG 6, below:



FIG 6: *An example of the message posted on reddit health forums.*

I received 27 replies indicating a willingness to be interviewed (plus four polite declines), and from these, 13 interviews were eventually conducted successfully. The reasons for not being able to interview the 14 other volunteers were being unable to arrange an appropriate time (10), or concerns over privacy (3); one stopped communicating with no reason given. In one case, I did not conduct a full interview as such, but had several private

message exchanges with a moderator [coded as RM010], and some quotes from this moderator are included in Chapter 6 along with those who were fully interviewed. Three additional interviews were carried out with moderators I was referred on to, and with whom I then corresponded with directly. I also interviewed one moderator of a non-reddit health forum – the LupusUK forum hosted by HealthUnlocked, a platform hosting discussion forums on behalf of health charities – as a check to see if a non-reddit moderator's experiences seemed significantly different (they did not). Eighteen interviews were undertaken between December 2015 and April 2016. Throughout the rest of this study, the identity of the participants is protected by coding them as RM (Reddit Moderator) 001-018, in the order in which they were interviewed.

Moderators were self-selecting for interview and as such may be subject to bias. Sending a personal message to the moderation team, and sending a personal message to individual moderators were equally successful recruitment methods, each accounting for seven interviews. The only moderator who replied positively to open posts I made on their forums was the one from HealthUnlocked, but three reddit moderators replied to say that surveys were not encouraged. All three moderators I was specifically referred onto did agree to be interviewed. The moderators were interviewed by Skype where they were agreeable to this, and by email or through the reddit internal messaging system if they were not; 10 moderators agreed to a Skype interview with those that declined citing privacy concerns. Skype interviews were audio recorded using Call Recorder for Skype software, and ranged in length from 25 minutes 53 seconds to one hour, 43 minutes and 18 seconds (average 47.5 minutes). The moderators who were not comfortable agreeing to a Skype interview were asked the same questions via the internal reddit private messaging (PM) system. Questions were emailed one at a time, so that the exchange mirrored a conversation, with options for deviation from the set questions or a request for further information as the interview progressed. In

the case of the both the Skype and PM interviews, the interviews were written up and sent to the moderators to check/amend, and to add any further information they felt was necessary. Of the moderators who provided personal information, 14 were male, 3 were female (one declined to answer). They ranged in age from 17 to “over 50”. Thirteen were U.S.-based, two were in the UK, with the remaining three in Fiji, Australia and the Dominican Republic.

I began by intending to interview only moderators of health forums, including those related to the two PHEICs which had occurred during the period of my study: r/ebola and r/zika, but based on suggestions from the moderators, I also expanded this to include r/science (on which there had been considerable activity related to Ebola as well as other health and medical topics) and r/starwars, which was suggested as a good example of a subreddit which had seen a dramatic increase in the volume and flow of traffic over a short period of time (coinciding with the release of *Star Wars VII: The Force Awakens*, the first new live action Star Wars movie for 10 years, in December 2015), similar to the dramatic increase in traffic that had been experienced on r/ebola and which might be experienced in the event of a future sudden disease outbreak.

During the process of interviewing moderators and gaining their trust, I was offered moderator status of three subreddits (r/zika, r/ebola and r/science). This afforded me access to additional information on moderator activity, including the moderation logs, in which all actions undertaken by moderators are listed, and ‘modmail’, private forums that sit alongside the subreddits where private conversations on issues relating to the running of the subreddit take place, and where messages sent to the moderators by users of the forum are received. Though the moderators were happy for me to take on an active role, after discussions with them I felt happier taking a backseat and simply observing their actions, to ensure that my participation did not influence the evolution or character of the subreddits in any way.

In addition to modmail, r/science moderators have access to r/ArmyofScience, a subreddit for the more than 1600 ‘comment moderators’ of r/science, who have privileges to remove comments that violate the forum’s strict rules. A key role of r/ArmyofScience is to enable the large number of r/science moderators to discuss issues relating to the forum, such as how to deal with troublesome posters and how to plan/prepare for periods when activity is expected to increase significantly, for instance around AMAs (question and answer sessions with experts) that are expected to be particularly popular. This provided additional insights into the moderator processes operating on reddit.

The interviews I conducted, and the access to the moderation logs I was granted, provided valuable insights into how reddit moderators influence and control the quality of information that appears on the site. The 18 moderators I communicated with came from 14 different forums between them.

	MODS INTERVIEWED	NUMBER OF MODS	SUBSCRIBERS
Cystic fibrosis	1	7	1,110
Chronic pain	1	3	7,746
Chronic fatigue	1	1	2,499
Fibromyalgia	1	5	4,256
Lupus (not reddit)	1	2	Not available
Bipolar	1	7	17,794
Ebola	5	8	9,359
Public health	1	3	7,242
Zika	3	2	129
Zika Virus	1	8	348
Rare diseases	1	5	102
Star Wars	2	18	471,921
Science	1	1,181	11,832,419
Theory of Reddit	1	10	57,597
		Median: 6.0	

TABLE 4: *Forums included in this study and the number of moderators on each. The total number of moderators (21) appears greater than the number interviewed (18) as some moderated more than one of the forums shown.*

Eleven moderated more than one subreddit, and the subreddits ranged in size from 102 subscribers (for r/rare diseases) to 11,832,419 (for r/science) at the time the interviews took place. All but one of the subreddits (r/chronic fatigue) had a team of moderators, with the size of the moderation team ranging from two to 1,181 (for r/science). The median number of moderators per forum was six.

### 3.5 Summary of data collected

	QUALITATIVE	QUANTITATIVE
Phase I	Interviews with 14 NGO workers	
Phase II-A	Responses in free text boxes on survey	17 survey responses providing 393 data points
Phase II-B	Responses in free text boxes on survey	27 survey responses, providing 1,140 data points
Phase III	[1] Content of posts on r/ebola [2] Content of discussions in modmail forum [3] Interviews with 18 moderators	Logs of moderator actions Posts on r/ebola

TABLE 5: *Qualitative and quantitative data collected and analysed.*

## 4: REDDIT – ‘THE FRONT PAGE OF THE INTERNET’

---

### 4.1 Introduction

In this chapter I describe reddit, the platform that hosts the r/ebola forum I selected for my case study. I explore its structure and characteristics, and consider what advantages or disadvantages these may offer health information seekers during a public health emergency.

Reddit ([www.reddit.com](http://www.reddit.com)) is one of the world’s most popular websites. Sources differ slightly in exactly how popular (reddit stopped displaying user statistics following a redesign in May 2016) but reported figures suggest this is high and growing. Choi et al (2015) report it had 169 million unique visitors in May 2015; Morse (2015) reports 172 million in June 2015; figures of 231 million (Smith, 2016) and 243 million (statista.com, 2016) are reported a year later. Figures for account users remain more constant, from 36 million in June 2015 (Morse, 2015) to similar numbers, in 217 countries, in November 2016 (Smith, 2016). There is no doubt that reddit represents a remarkably large crowd. If it was a country, its number of active account users would be roughly the same as the working population of the UK. As an information source, it has many more unique visitors than the most-read professional online newspaper, *The Daily Mail*, which had just under 14 million unique readers in January 2015 (Baird, 2015). Digg, a similar aggregator platform, had less than 12 million monthly active users (Digg, 2015) at the same time.

Demographics of reddit users also vary depending on which survey is consulted, but they seem more likely to be male than female (67:33 according to Pew, 2016; 63:33 according to reddit (Reddit, 2016), which also offers a third option of ‘other’, chosen by 3.5%); young (64% aged between 18 and 29, Pew, 2016; 56.7% between 18 and 25, according to reddit’s own survey, compared with 22% of the U.S. population in the same age group); to have a

college-level education (42% compared with 28% of the general population, but this may be a factor of its younger overall age); and to self-identify as liberal (43% compared with 24% of the general U.S. population, Barthel et al, 2016). In total, 55% of reddit's users (according to its own survey) are based in the U.S., with the next largest countries Canada (8%) and the UK (6%). Bearing in mind the different population sizes, while approximately 4% of adults use reddit, the proportion for the UK is higher, with reddit reporting 6.6 million users from the UK (10% of UK population) per month in mid-2015 (Morse, 2015). It is important to acknowledge, however, that even though it is widely used, the observed and anticipated behaviour of the reddit userbase may not be generalizable to the population of any country or region, or in fact any other online community.

In technical terms, reddit is a web-based platform which collects, or 'aggregates' content from across the worldwide web. Users can read content submitted by others and those who register with the site can also contribute content themselves, either by posting links to external content or by generating content hosted on reddit. Submitted content can be voted on by registered users and is ordered by an algorithm in which the votes users have cast have a strong influence on how prominently posts are displayed. Users can also make comments on the submitted content; these too can be voted on by other users with the votes influencing the order. In keeping with its history and original intention of being a news site, there is also a strong temporal element to how information is ordered. The most prominent posts – of which the top 25 appear on the Front Page – are selected using an algorithm known as the reddit 'hotscore', a value based on the time the submission was made, to which additional value is added or subtracted depending on how the submission is voted on. The most prominently displayed information has been posted recently and voted on favourably by the community.

First, it is useful to set out some terms that will be used throughout this thesis and how they will be defined.

**Front Page:** This refers to the succession of posts a reddit user first sees when they enter the URL [www.reddit.com](http://www.reddit.com) in a web browser or click on the reddit app. The Front Page is, essentially, a ‘shop window’ for everything reddit contains. Each user sees a different version of the Front Page depending on which parts of the site they have subscribed to, with the exact order of what is displayed determined by the votes cast by the reddit community.

**Subreddits:** Content on reddit is ordered into subreddits, subsections of the site arranged around specific topics. When users register with reddit they are automatically subscribed to [r/popular](https://www.reddit.com/r/popular/), a subreddit that mirrors the Front Page but removes pornographic and adult material, and can choose to subscribe to other subreddits at any time according to their interests. Users can enter reddit and interact with its content through the Front Page, or through individual subreddits. Each subreddit has its own front page, on which content is ordered according to the ‘hotscore’, and the content on the main Front Page is selected from the most popular content across all the subreddits.

**Post:** This refers to a submission made to reddit – the content that is ‘posted’ to the website. The content itself may be hosted on an external website or on a subreddit. A post is displayed as hyperlinked text which, when clicked on, brings up the actual content submitted. The person who makes the post is referred to as the ‘poster’.

**Comment:** This refers to a submission made in response to a post – which may be an answer to a question asked in the original post, a comment made on a statement or news story, a personal anecdote offered in support of the original poster, or an additional link to information that supports or refutes the original post. All comments are made in response to a specific post, and always remain linked to that post. They are displayed beneath the original post

(sometimes abbreviated to OP) in hierarchical comment trees, which can present long and involved arguments.

**Votes:** Post and comments both receive votes from reddit's registered users. They can receive 'upvotes', which indicate they have been received positively, or 'downvotes', which indicate they have been received negatively. The more favourably posts are received, the more prominently they are displayed.

In terms of its CIG, reddit comprises several *Collections*:

- It is a *Collection* of submitted content, on a variety of topics, including posts and comments.
- It is a *Collection* of subreddits, more specialised topic areas which help users to find content relevant to their interests.
- It is a *Collection* of registered reddit users – known as 'redditors'.

Kassing et al (2015) highlighted an important distinction between two of these *Collections*: 'candidate knowledge items', or pieces of information submitted in posts and comments which are visible on the site and 'candidate knowledge experts', the users (and subreddit communities) to which reddit gives access, who may hold additional information but have not yet deposited it.

The Front Page, in contrast, is a *Collaboration* between many reddit users who *Decide by Voting* which of the items of content are most worthy of other users' attention and therefore displayed most prominently.

The CIG of reddit is therefore rather complex, and while the user community collectively curates the content, does it do this intelligently? In a public health emergency, there may be a plethora of information but people may have neither the time to read everything nor be sure which sources to trust. It is therefore valuable to determine if reddit can provide a valuable shortcut to useful information available across the Internet.

## 4.2 Structure of reddit

Reddit can be accessed through any web browser using the URL [www.reddit.com](http://www.reddit.com) on desktop or mobile platforms. This presents the user with the Front Page, on which the most highly valued posts are displayed, and which has several features that enable interaction with the site and its content.

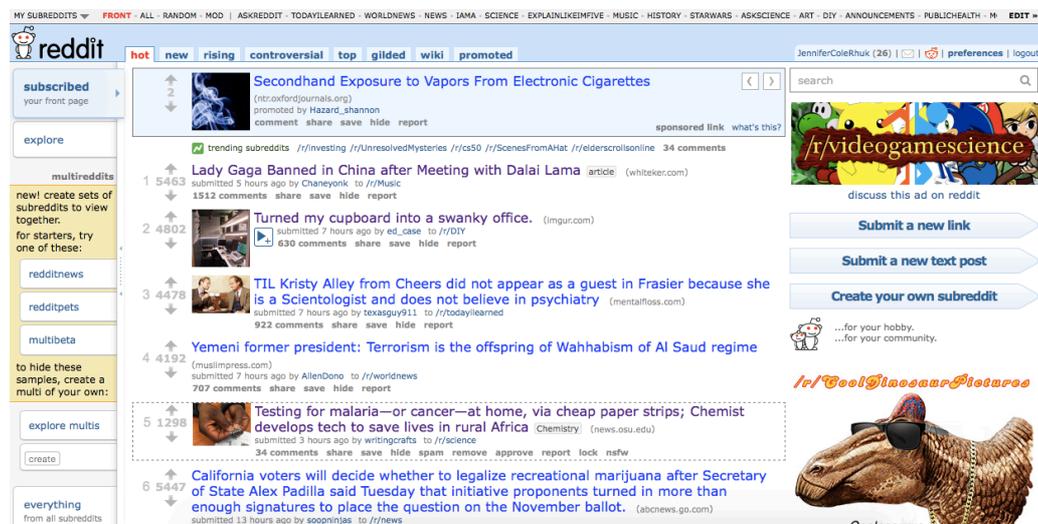


FIG 7: A typical example of a reddit Front Page (screenshot taken at 16:05 on Wednesday 29th June 2016) showing posts, sponsored posts and adverts.

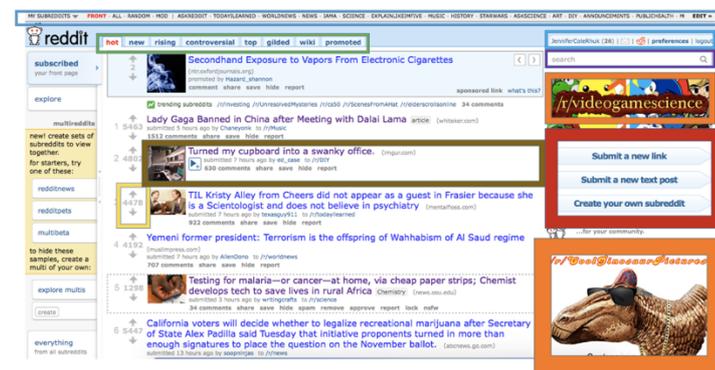
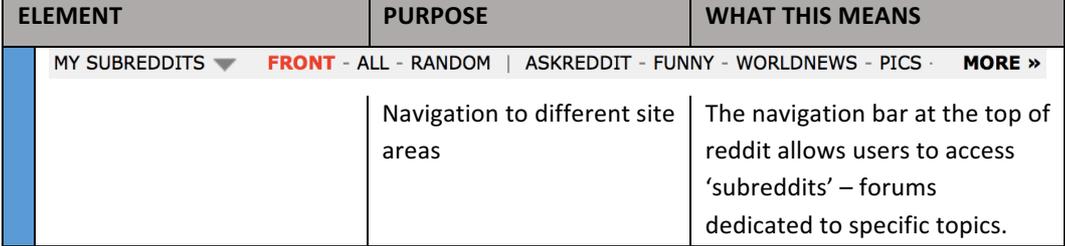
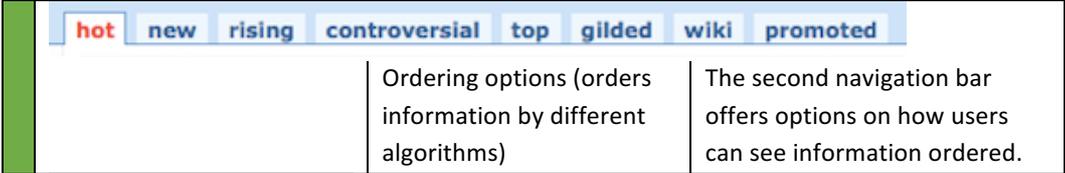
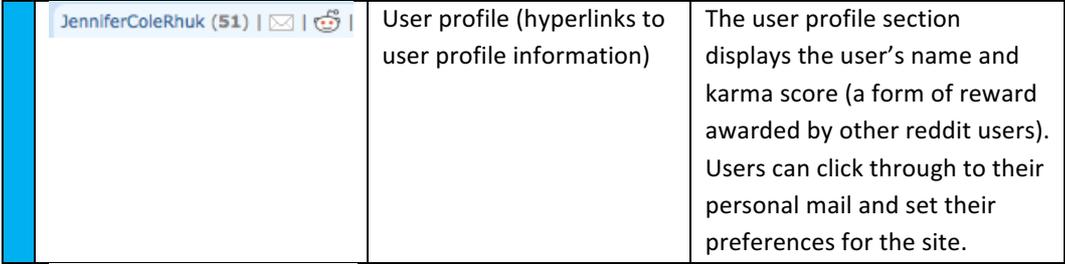
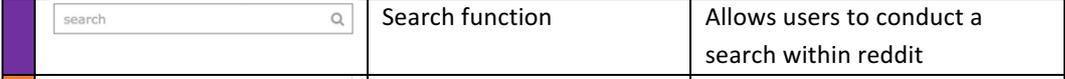
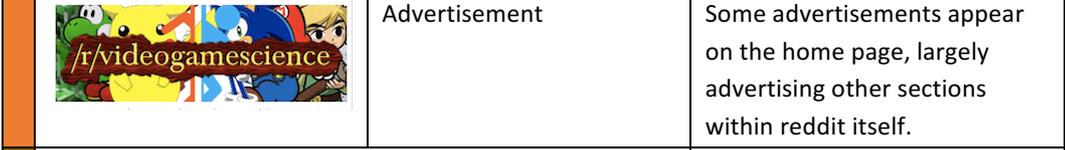
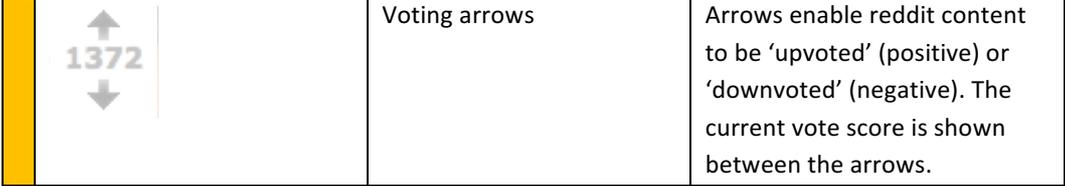
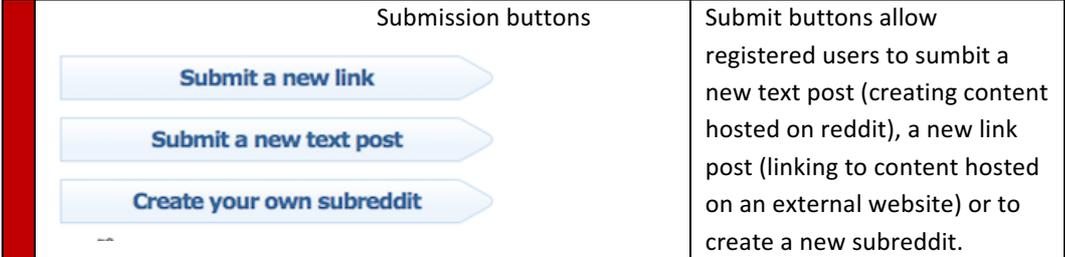


FIG 8: The Front Page contains several different elements, such as links to content, voting buttons, ordering options

and advertisements. Each of these is contained within a different coloured border on this figure. The function and purpose of each is described in TABLE 6 on the following page.

TABLE 6: Elements of the reddit Front Page

ELEMENT	PURPOSE	WHAT THIS MEANS
	Navigation to different site areas	The navigation bar at the top of reddit allows users to access 'subreddits' – forums dedicated to specific topics.
	Ordering options (orders information by different algorithms)	The second navigation bar offers options on how users can see information ordered.
	User profile (hyperlinks to user profile information)	The user profile section displays the user's name and karma score (a form of reward awarded by other reddit users). Users can click through to their personal mail and set their preferences for the site.
	Search function	Allows users to conduct a search within reddit
	Advertisement	Some advertisements appear on the home page, largely advertising other sections within reddit itself.
	Post	This the main content of the Front Page. The title hyperlinks to content hosted on a subreddit or external website
	Voting arrows	Arrows enable reddit content to be 'upvoted' (positive) or 'downvoted' (negative). The current vote score is shown between the arrows.
	Submission buttons	Submit buttons allow registered users to submit a new text post (creating content hosted on reddit), a new link post (linking to content hosted on an external website) or to create a new subreddit.

Being able to use all these features depends on whether the user is registered (signed up for an account) with reddit. To register, users do not have to surrender any personal information and do not even have to provide an email address. Registration requires them only to choose a username and password, which allows them to log in to that username account. Registered users can make submissions and vote on other posters' submissions, while unregistered users can read posts and comments but cannot add or vote on content.

Reddit therefore provides users with several options for interaction, including reading and submitting information, debating contentious issues, and showing interest and appreciation in information others have posted (Hsieh et al, 2013; Kassing et al, 2015). These include:

- Passive reader ('lurker')
- Content creator (original poster)
- Content commenter (poster)
- Content voter

The percentage of reddit users who contribute and interact with the site broadly follows the 90-9-1 rule of the Internet (Nielsen, 2006, Choi et al 2015), with 90% interacting passively (just reading), 9% having some interaction (posting and voting) but only 1% interacting regularly. This can vary across subreddits, however. For instance, according to a discussion on r/theoryofreddit in 2016<sup>5</sup>, r/askreddit had 15 million unique views and 500,000 unique contributors; r/pics and r/funny had 10 million unique views and 200,000 unique contributors; r/worldnews and r/news have 6 million unique views and 100,000 unique contributors. Figures do not necessarily

---

<sup>5</sup> [https://www.reddit.com/r/TheoryOfReddit/comments/3qbbsp/do\\_we\\_know\\_the\\_percentage\\_of\\_redditors\\_who/](https://www.reddit.com/r/TheoryOfReddit/comments/3qbbsp/do_we_know_the_percentage_of_redditors_who/), last accessed 10<sup>th</sup> March 2017

remain static even on one subreddit: at the height of interest in r/ebola, in mid-October 2014, there were on average around 70,000 unique visitors per day, and 250 posts (0.0036 active posters for every unique user, but some of the non-posters may have been voting or commenting, for which figures are not available), but earlier in its history, at the beginning of August 2014, there were approximately 500 unique visitors per day, who made on average around 20 posts per day (0.4 posters for each unique visitor). The ratio can change depending on which subreddit is considered, and at what time; some of the factors affecting this are explored in Chapters 6 and 8.

### *A collection of posts*

The main purpose of reddit is to display ‘posts’ – items of content – in a rolling news format, with 8-10 posts generally being visible on a typical computer or tablet screen at any one time; users can scroll down 25 posts before being asked to click through to a second page. Posts are constantly replaced by newer, more timely ones. Originally, posts could only link to content hosted on external websites – reddit users ‘collected’ links and the reddit platform aggregated them in one place – but increasingly posts link to content hosted on reddit itself (Merritt, 2012, Singer et al, 2014).

The Front Page indicates which, of out more than 200,000 posts made on average by reddit users every day (Smith, 2016), have been received most favourably by the reddit user community. Each post title can be clicked on to take the reddit user to the content to which it refers – such as, in the case of the posts shown in FIG 7, a news item hosted on the entertainment website [www.whiteker.com](http://www.whiteker.com) about the pop star Lady Gaga’s recent meeting with the Dalai Lama [post 1], a photo of a reddit user’s DIY project to build a home office hosted on the external image hosting website [imgur.com](http://imgur.com) [post 2], or a

report on recent cancer research carried out at the University of Ohio hosted on the university's own website [post 5].

There are two basic types of post that can be made on reddit:

- **Link post**, which links to content hosted on external websites.
- **Self post**, which usually contains text such as a question, statement or anecdote (with a generous 10,000 character limit) and does not link to external content. Its content sits in the subreddit to which it is posted.



FIG 9: A series of link posts from the subreddit r/zika\_en. Clicking on the title of the posts takes the user to external content. The original source (e.g. CDC, WHO, New York Times) is indicated at the end of the post title.



FIG 10: A self-post asking for advice regarding the Zika virus. Clicking on the post title brings up text which, in this case, asks a question. Other reddit users can then answer or comment, and other users can agree, disagree or post a different reply.

TABLE 7: *Elements of a post*



ELEMENT	PURPOSE	WHAT THIS MEANS
	Upvoting and downvoting arrows (number between them shows the positive:negative vote ratio).	Clicking on the 'up' arrow enables a registered reddit user to cast a positive vote; clicking on the 'down' arrow casts a negative vote.
	Image	Posts may have a thumbnail image giving a clue to their content, or could be the content itself; some posts link to art images or photographs of DIY projects.
	Subject flair	Posts may have 'flair', which gives an indication of subject matter (e.g. chemistry) or source.
	Source indication	Indicates and links to the source of the content: an external website, or, if a self-post, to its subreddit.
	Time indicator	Posts are time stamped with when they were originally submitted.
	Poster username (with hyperlink to user home page)	Posts indicate the username of whoever submitted it and whether they originated its content.
	Subreddit indicator (with hyperlink to that subreddit)	Posts are submitted to subreddits, with the most popular reaching the Front Page. The subreddit to which the post was made is given.
	Number of comments (with hyperlink to comments)	The number of comments made on a post is indicated. A user can click on this to read the comments or add their own.
	Possible user actions	Users can act on a post in several ways including sharing it on other social media, reporting it if they feel it is inappropriate, or hiding it from their Front Page.

Key ways in which reddit users can interact with posts (self or link posts) include leaving a comment on the post, which other users will be able to see by clicking on the 'comment' tab under the post, or by voting favourably ('upvoting') or unfavourably ('downvoting') on the post by clicking on the arrows found to the left-hand side of the title. A user can only upvote or downvote a single post once: a second vote by the same user cancels out the previous one, preventing the same person from voting multiple times.

Each post contains elements that add to or enhance its function. In addition to the title, which links to the main content, it contains information indicating where the content originally appeared (on an external website or in a subreddit), the username of the reddit user who posted it, the subreddit to which it was originally posted, how long ago it was posted and the ratio of positive to negative votes that have been cast on it. There is also a series of actions that can be taken – such as 'share', which enables the post to be shared on social media such as Facebook and Twitter, or 'Report', which enables other users to report inappropriate posts to the reddit administrators.

When submitting a post, reddit users can choose whether to submit to a single subreddit (for example, a question about a child's rash could be posted in r/parenting, r/askdocs or r/infectious disease), or to post in more than one subreddit simultaneously, known as cross-posting. The positioning of the subreddit chosen is likely to influence how submissions are received: for instance, a news story suggesting the cause of the Zika virus is genetically-modified mosquitoes may be received much more favourably, and receive many more upvotes, in r/GMOinfo – a subreddit in which people discuss their concerns over genetic modification– than in r/publichealth or r/science. If the intending poster feels no subreddit is suitable for the post they want to make, they may choose to create a new one (Rivera, 2016).

It is useful to consider which subreddit posts were submitted to when deciding how strongly to trust them. This influences the message credibility layer of the Wathen & Burkell model, and I will return to this in Chapter 6.

### *Comments on reddit posts*

Once a user has made an initial post, subsequent users may comment on it, or comment on previous comments that have already been made, enabling topics to ‘branch out’ in hierarchical comment trees (Weninger et al 2013; Choi et al, 2015). The number of comments made on a post is indicated beneath it (e.g. 36 comments in the example given in TABLE 7). Clicking on the word ‘comments’ brings up those comments for the user to read and branches can be collapsed by clicking [-] next to the user name, so that surrounding top-level comments can be seen more easily.

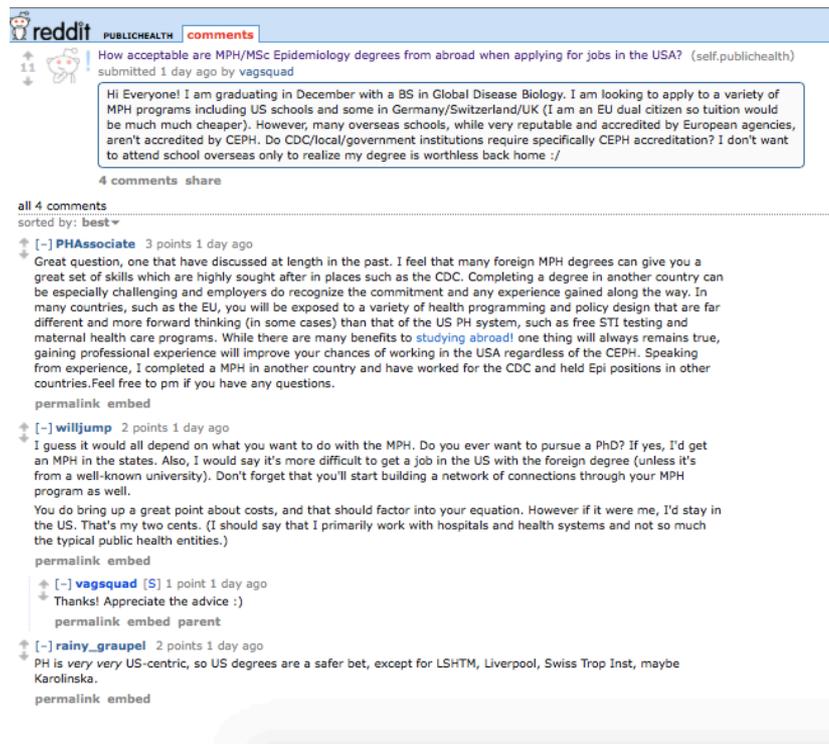


FIG 11: *Reddit registered users can comment on posts, with comments displayed in threaded hierarchical comment trees.*

Every post and comment which has been made in reddit's history (bar those which have been removed for breaching site rules or by the person who originally made them) are still available to read. Six months after a post is submitted, the thread (the post and its comment trees) is automatically archived, meaning that users can no longer comment or vote on it, or vote on the comments it attracted, though it remains visible and can be searched for.

This structure is generally considered to create a well-ordered, well-formed and permanent discussion platform that enables divergent topics to emerge and encourages robust discussion (Weninger et al, 2013), though this view is not universal. Others view reddit as messy, and suggest that a better ordering system might be to cluster posts based on opinion (Gao, 2016) but in my view, such an alternative system would be more likely to create the echo chamber effect for which social media is often criticised, and would risk diluting reddit's value as a place where diverse opinions can be expressed and debated.

### 4.3 Voting on reddit

Posts can also receive votes. All registered users on reddit have the option to 'upvote' or 'downvote' any content they read by clicking on the up or down arrow found to the left-hand side of post titles. The voting score (the ratio of upvotes to downvotes) is indicated between the arrows. Voting gives a general indication of how well, or not, a post or comment has been received by the community (Bross et al, 2012).

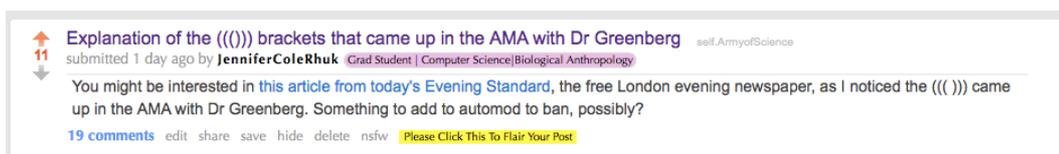


FIG 12: *Votes are cast using the up or down arrow to the left-hand side of a post, and play a part in determining how prominently posts are displayed.*

Reddit's voting structure is unusual amongst peer-to-peer platforms by giving users the option of downvoting content – thereby affording it a negative score – as well as a positive one. Facebook for example, currently only allows users to 'Like' content, not to 'Dislike' (though they can choose to be 'sad' or 'angry', suggesting some level of disapproval with the presented content).

### ***Voting and 'karma'***

As well as determining the position and prominence of posts and comments, votes cast also go towards the user's 'karma score', indicating that the user has submitted content considered to be of value by the reddit community. The karma score of the poster does not influence how their posts and comments are displayed, however: if a poster has 1000 karma points, this has no greater influence on how a post they make is positioned than a submission made by a poster with 0 or -1000 points, though accumulation of karma brings some benefits to the user, such as being able to set up one's own subreddit, which is dependent on having gained at least 100 karma points.



FIG 13: *Clicking on the username of the poster brings up their homepage, which shows how much karma they have been awarded by other users.*

The poster's karma score is not routinely displayed on the post or comment, but is shown on their user home page. While this can be viewed by any registered user by clicking on the username displayed by the post, a conscious decision to check it out needs to be made. There are in fact two separate karma scores – one for post karma, which indicates votes cast against the user's posts (both link posts and self-posts), and one for comment karma, which indicates karma gained for votes on the user's comments. Clicking on the 'show karma

breakdown by subreddit' link below this reveals which subreddit the post and comments that have attracted the karma were cast in. Karma scores can be used by other redditors to check whether a user has a positive or negative posting history, and which subreddit the positive votes they have received were cast in. If a poster has positive karma but has mostly accrued this in a subreddit known to support racist views or conspiracy theories, for example, this may be viewed as an indication of negative activity or disagreeable views by other redditors, in the same way as a negative karma score.

#### **4.4 Ordering of information on reddit**

Reddit has developed several ways to order the information it presents. First, by breaking itself down into a series of smaller communities known as subreddits, information on specific topics can be found more easily. Second, content is arranged into a specific order that displays some posts more prominently than others.

'Subreddits', the topic-specific forums into which posts are grouped, became necessary as reddit grew too big to stay as a single community – the original r/reddit. Each subreddit is run by a moderator, or team of moderators, who are generally not employees of reddit but interested active users – volunteers. They set their own rules about what can be posted to their subreddit and, in some cases, determine who can post.

Subreddits range in size, purpose, social norms, and levels of activity (Leavitt, 2015). The largest ones have millions of subscribers, such as r/IAMA (16,315,870 in April 2017), r/science (16,539,470 in April 2017) and r/news (13,986,360 in May 2016) to just a handful in others. In 2014, 212 out of more than 300,000 subreddits had more than 100,000 subscribers (Jiang et al, 2014).

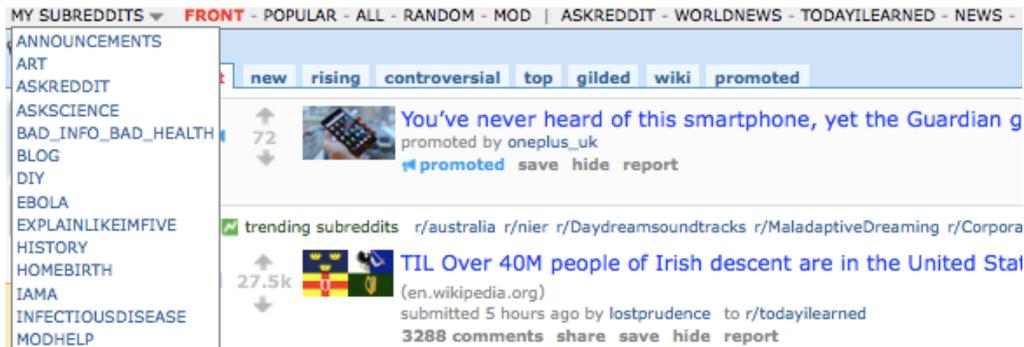


FIG 14: Users can click through to subreddits using the menu bar at the top of the page, or from a dropdown menu accessed on the left-hand side.

Until February 2017, when new users signed up to reddit they were automatically subscribed to approximately 50 ‘default’ subreddits, and could choose to subscribe to others depending on their interests; they were free to unsubscribe from some or all the default subreddits if they chose. Since February 2017, this has changed and now new users are signed up to r/popular, a subreddit that aggregates posts in the same way as the ‘old’ home page, but removes pornographic subreddits, those that actively opt out, and subreddits that many reddit users consistently filter out. Users joining since February 2017 can still subscribe to additional subreddits, whose content will appear mixed in with the r/popular posts. For existing users, their home page remained the same, as if they had actively chosen to subscribe to the defaults.

Subreddits to which a user is subscribed appear in the menu bar on the Front Page, enabling them to click through, and sufficiently high scoring posts from their selected subreddits will appear amongst those from the default subreddits or r/popular on their Front Page. Thus, every registered user sees a personalised version of the Front Page depending on which subreddits they have subscribed to (Weninger et al, 2015). Unregistered users see only posts from r/popular. Once in a subreddit, the front page for that subreddit shows only posts that were made specifically to it. Most subreddits are open,

enabling any registered user to subscribe (by clicking on a button that appears on the subreddit's own front page), but others are 'restricted' (all users can see information but only users subscribed to that subreddit can contribute content or vote), or 'private' (only subscribers can see information). Restricted and private subreddits can only be subscribed to by invitation – usually from the moderator(s); a private message is sent to the invitee with a link, through which they can accept or decline by return. There are subreddits on virtually any topic, from sports to crafts, to politics, which I chose to categorise into five broad types:

TYPE	PURPOSE	EXAMPLES
<b>Informative I:</b> News aggregators	1. To collect news stories from different sources for information 2. To enable discussion of news stories and events	r/news, r/worldnews, r/movies, r/soccer
<b>Informative II:</b> Question and Answer	1. To enable posters to ask questions, which can then be answered by other users or experts	r/askscience, r/askdocs, r/explainlikeimfive, r/AMA
<b>Entertaining:</b> Funny/entertaining	1. To enable posters to share funny and entertaining images and stories	r/funny, r/pics, r/jokes
<b>Communities I:</b> Support	1. To offer support, help and advice to community members	r/cancer, r/depression, r/parenting
<b>Communities II:</b> Special interest	1. To enable people interested in a topic to discuss it with others	r/starwars, r/diy, r/books

TABLE 8: *Categorisation of subreddit types*

Each subreddit is, in emergency management terms, a Community of Interest (Australian Emergency Management Institute, 2013). It comprises a group of people who do not (necessarily) share a geographical location, but who have chosen to seek out information on a single topic – in the case of the health forums I am interested in, an infectious disease – such as Ebola or Zika, or a health condition such as diabetes or cystic fibrosis. They may simply want information, provided by the informative subreddits, or to connect with others who share specific interests, provided by the community subreddits.

In terms of the above classification, most health forums on reddit are a deviation from its original purpose and still the main expression of its Collective Intelligence Genome, which is to be a *Collection* (of news stories), best described by the category Informative I: News Aggregators, towards being *Create* – the creation of a supportive community as described by the categorisation Community II: Support. And yet, when one looks at r/ebola and r/zika they display characteristics more common to informative forums than to communities, such as higher percentage of link posts than self-posts and a higher proportion of passive readers than active posters (Singer et al, 2014; Kassing et al, 2015); they are collections of news stories containing link posts that generate little discussion or long comment threads. Though r/ebola did edge towards characteristics more common to support forums at the height of the outbreak – as I will explore in Chapters 6 and 8 – even at that time it still appears to have been predominantly used as a special interest community, not a support community. This is perhaps not too surprising, as outside of North America, Europe and India, use of reddit is negligible. Reddit users were not living with Ebola or in any real danger of catching it – they were interested in the outbreak from afar. In r/cancer, r/cystic fibrosis etc, most users are living with the condition themselves or have a close friend or relative who is. I return to the influence this had r/ebola in Chapter 5 and Chapter 8.

### ***Searching for new topics***

While the subreddits serve as a kind of library system for information, reddit also has an internal browser. If the user is interested in information on a specific topic and is not already subscribed to a subreddit on that topic (or aware if one exists), they can use the search box at the top of the right-hand corner of the page. This brings up a list of any subreddits dedicated to the search term, and all posts in which the search term appears. Subreddits and/or

individual posts are then accessed by clicking through from the results returned by the search.

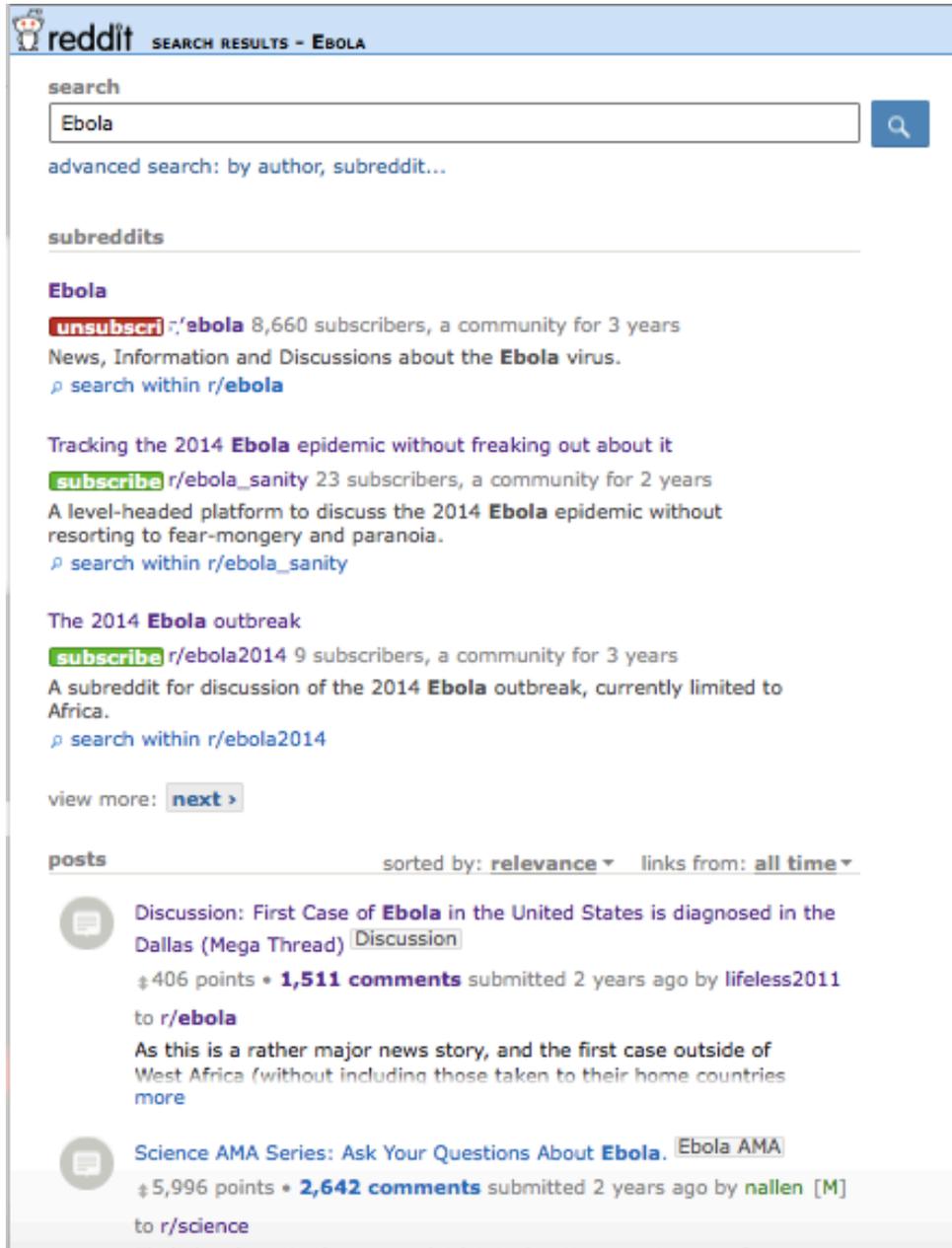


FIG 15: Search results page showing subreddits and posts returned on a search using the keyword 'Ebola' within reddit on 20<sup>th</sup> April, 2017.

## Ordering posts within reddit

The second way reddit orders information is determined by how its system is configured to display posts, on the Front Page and on the first page of the individual subreddits. Its unique ordering system is a key technological characteristic and is considered one of its strengths: the algorithm behind this appears to be main reason for Condé Nast and Advance Publications' interest in investing in the site during its early history (Wired, 2006). It is worth describing this system – known as the reddit ‘hotscore’ – in detail.

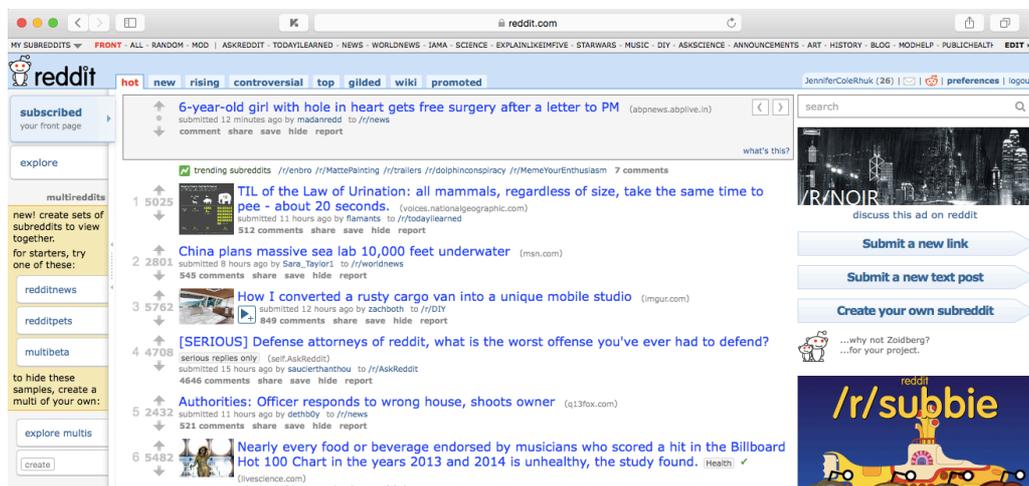


FIG 16: The order in which posts are displayed on reddit is determined by the hotscore algorithm. The post with the highest hotscore appears at the top.

Reddit is implemented in the programming language Python and its code is openly available at <https://github.com/reddit/reddit>. Posts and comments are ordered by a series of algorithms implemented in Pyrex, a language to write Python C extensions. There are two main components that determine the ordering mechanism of posts on reddit: the sum of upvote-downvotes the post has received, indicating how well it has been received by the reddit community, and how recently the post has been made, indicating its newness.

The equation for the hotscore,  $h$ , is computed as follows:

$$\begin{aligned}r &= \text{sgn}(s) * \log_{10}(\max(|s|, 1)) \\s &= 1 + \#\text{upvotes\_of\_post} - \#\text{downvotes\_of\_post} \\t &= (\text{date\_in\_seconds\_post\_was\_made} - \text{date\_in\_seconds\_reddit\_started})/45000 \\h &= r + t\end{aligned}$$

The hotscore is a ranking system, which ranks posts relative to others. For each post,  $(t)$  is the time in seconds that has passed since the hotscore was introduced (recorded as 07:46:43 on 8<sup>th</sup> December 2005, which has a Unix Time Stamp of 1134028003). Each individual post therefore has its own  $(t)$  value, which remains constant once it has been posted, while each subsequent post will have a different, larger  $(t)$  value as  $(t)$  is constantly increasing as time passes. If posts were ranked on  $(t)$  alone, they would simply be presented in the chronological order in which they were posted, with the most recent appearing first (and in fact, reddit users can choose to display them this way, by choosing the ordering option ‘new’ rather than ‘hot’).

The  $(t)$  value is then combined with *order* ( $r$ ), an additional value that, when positive, is added to  $(t)$  or, when negative, is taken away. A positive  $(r)$  ‘buys’ a post more value, thus pushing it above more recent posts, whilst a negative  $(r)$  value penalises it, pushing it down the rankings, as if it had been made later. The order in which posts are displayed is therefore influenced by  $(r)$ .

A component of  $(r)$  is the *score* ( $s$ ), determined by subtracting the number of downvotes a post receives from the number of upvotes it receives (i.e. 10 upvotes minus 2 downvotes will result in a score of +8). Due to the  $\log_{10}$  in the equation for  $(r)$ , the first 10 votes each make a significant addition to a post’s overall  $(r)$  value, but this diminishes as additional votes are cast and means that earlier votes add more value to the order than later ones. This is an important element of the algorithm, as without it, it would be difficult for new

stories to get a ‘foothold’ on their subreddits and to make it to the Front Page. There is no ‘threshold’ ( $r$ ) value for ensuring a post reaches the top of its subreddit or makes it to the Front Page; this will be determined by how many other posts it is competing with. Posts on a relatively small subreddit such as *r/rarediseases*, which has only 200 subscribers and around 10 posts a month, may need only 1 or 2 upvotes to appear on its subreddit front page, whereas posts on a more active subreddit, such as *r/news*, with millions of subscribers, is likely to need a positive score of more than 100.

A second feature influencing the order is that the ( $s$ ) value includes a +1: when a post is made on reddit, it is immediately awarded one upvote. Posts therefore begin from a score of +1 rather than 0. The rationale for this is that the poster will want to give their own submission a positive score and so the system does it for them. The first downvote reduces this to 0, the second to -1, and the third to -2. This affects how the posts are displayed, as the  $\text{Log}_{10}$  function requires an absolute value for the score  $|s|$ , and for this to have the (max) function attached to it to ensure that a positive value of at least +1 is produced (for which a  $\text{Log}_{10}$  value can be calculated). The sign of the score –  $\text{sgn}(s)$  – is then multiplied back in, creating a ‘blank’ space in the ( $r$ ) values between no votes and two negative votes being cast, shown in FIG 17.

The effect this has is that one or even two downvotes has very little effect on a comment’s score and position: at least three downvotes must be cast before a negative score registers. This protects posts against malicious downvoting. If two redditors are arguing, then each person alone cannot much affect the priority of the other’s posts or comments. For malicious downvoting to be effective, a group of at least three downvoters is required. This function seems to have been an accident to start with, but has since been deliberately kept because of its identified value.

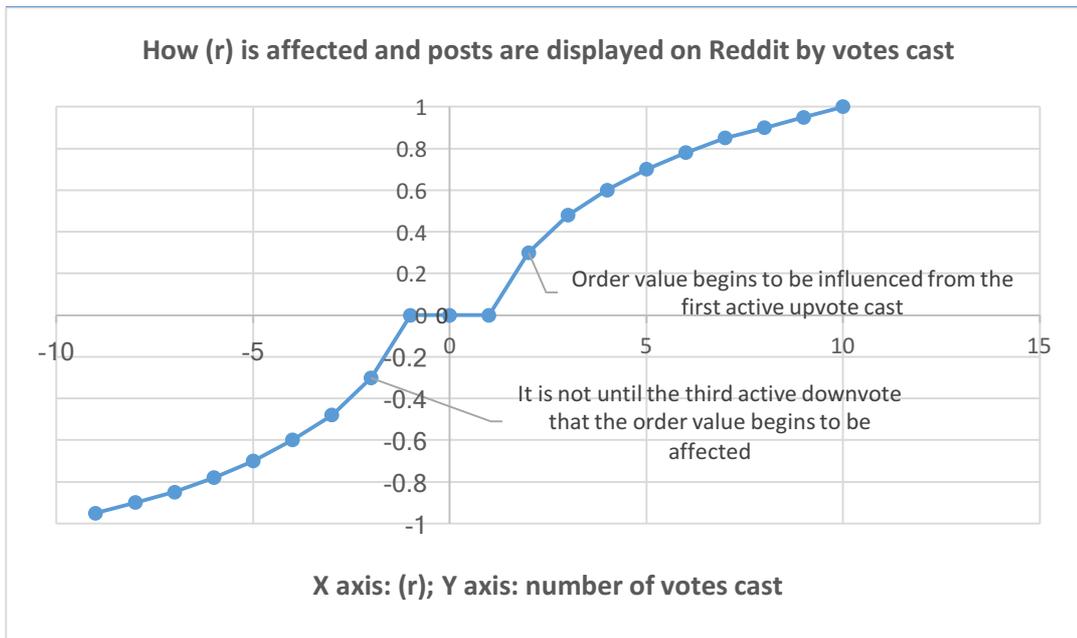


FIG 17: *(r)* affects the position of posts on reddit, dependent on votes cast

Originally, Reddit displayed the actual number of upvotes and downvotes against posts, but in 2013 the administrators began experiments to ‘fuzz’ vote totals as an anti-spamming measure. In July 2014, the old system was fully replaced by a ‘points system’ which shows the score based on the ratio of upvotes minus the number of downvotes, rather than actual numbers of votes cast. In addition, some subreddits – particularly the larger ones – hide the number of votes cast for a period after the post has been submitted, typically 60 minutes, so that other users cannot be influenced by early positive or negative voting (although, of course, a post’s position relative to others can give away a lot about how well, or not, it is scoring).

A post or comment which receives a significant number of negative votes is pushed down the order, and so is less prominent. Even if it is visible – usually because only a small number of submissions have been made to its subreddit, too few to knock it onto a second or subsequent page – it may be deemed

‘below threshold’ and its content hidden from view. This threshold is set by reddit at -4, but can be altered in a user’s preferences if they want to see such comments; moderators can also set this threshold higher or lower than the default for their subreddit.

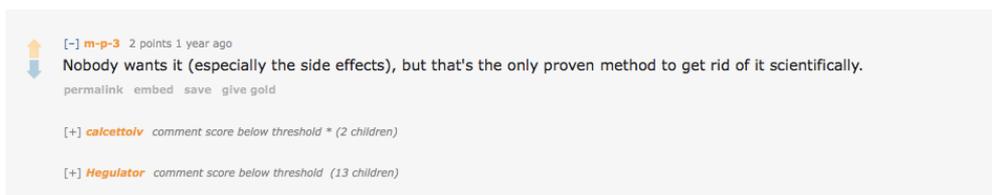


FIG 18: *Comments below a threshold are hidden, but can be opened by clicking on [+]* to reveal them.

Positive votes increase the score and essentially ‘buy’ posts extra time before they are overtaken by those made after: roughly, the first 10 votes ‘buys’ a post an additional ‘hot period’ of 12.5 hours (a period chosen arbitrarily by reddit, and which accounts for the 45,000 – seconds – in the algorithm). A negative score, giving a negative value to (r), subtracts from the time value and effectively sends the post back in time, pushing it down the hot-score-determined order until it is buried amongst older posts.

The voting system is a clear sign of community approval or disapproval that can help signpost users to content that is deemed to have more value, and away from content deemed to have less; this too influences message credibility in the Wathen and Burkell model. I will return to it in Chapter 6.

Reddit is a newsfeed and therefore places the greatest value on newness: its purpose is to provide *new* information to its users, rather than information *per se* and not only is the (t) value is considerably greater than (r), but also the overall influence (r) can have over time is limited by the  $\text{Log}_{10}$  function, as this means that every 12.5 hours, there would need to be 10-fold increase in

the number of upvotes-downvotes to maintain hotness. Few posts are likely to retain a high enough hotscore to remain on the Front Page after 50 hours, or four ‘hot periods’: it is extremely difficult for a comment to stay ‘hot’ for days. As time passes, even relatively high-scoring posts lose their ‘hotness’ relative to new posts and drop down the order, eventually disappearing from view all together.

### *Alternate ordering systems on reddit*

Posts on the reddit Front Page, and each subreddit, are ordered according to the ‘hotscore’ by default, but the user can choose other ways to see the information ordered, by clicking on one of the options at the top of the page.



FIG 19: *Tabs at the top of the page allow the user to decide how they would like to see posts ordered. Hot – ordered by the hotscore – is the default setting.*

The alternate options for ordering are as follows:

- **top** displays posts with the highest upvotes-downvotes score from a choice of time periods: past 24 hours; past week; past month; past year; all time.
- **new** gives preference to the most recent posts in order of time posted, regardless of what votes, if any, have been cast against them; when ordered by ‘new’, the top 25 posts on the Front Page will generally have been submitted within the last five minutes.
- **rising** displays posts that are attracting interest most rapidly, calculated according the number of times the link has been clicked as well as the number of votes it has attracted.

- **controversial** is determined by the ratio of upvotes:downvotes, rather than the score of upvotes-downvotes, indicating the posts that have caused the most controversy. It can be configured in different ways to display the most controversial posts from throughout the entire history of reddit, or from the past hour, 24 hours, week or month.
- **gilded** lists comments which have been ‘gilded’ or rewarded by users with reddit ‘gold’, a small payment that can be paid into one user’s account by another to show appreciation for the comment or post and buys access to reddit’s premium features. Choosing the ‘gilded’ ordering shows only gilded comments, but within this, the comments are ordered by hotscore.
- **promoted** displays sponsored links only.

### *Ordering comments on reddit*

While the default setting for posts is ‘hot’, since 2009 the default ordering for comments made on those posts has been a different algorithm, ‘best’, which does not contain the time component. This ensures that the most valued comments reach, and remain at, the top of a threaded comment tree under the associated post, regardless of when the comment was posted. Users can also choose to order comments by ‘top’, ‘new’, ‘controversial’, as well as ‘old’ or ‘Q&A’, the latter two features being unique to comments, but not by ‘hot’, while posts cannot be ordered by ‘best’. A dropdown menu at the top of the comment tree allows users to choose from the alternate ordering systems.



FIG 20: *Comments are default ordered by ‘best’, which has no time component.*

The algorithm for ‘best’ (shown below) also includes the lower bound of the Wilson score interval:

$$\left( \hat{p} + \frac{z_{\alpha/2}^2}{2n} \pm z_{\alpha/2} \sqrt{[\hat{p}(1 - \hat{p}) + z_{\alpha/2}^2/4n]/n} \right) / (1 + z_{\alpha/2}^2/n).$$

Where:

- $\hat{p}$  is the observed fraction of positive ratings
- $n$  is the total number of ratings
- $z_{\alpha/2}$  is the  $(1-\alpha/2)$  quantile of the standard normal distribution

This enables ‘best’ to order comments according not only to the ratio of upvotes:downvotes that have already been cast (as is used in the ‘top’ ordering) but, early in the voting process when only a small number of votes have been cast, the ratio they will likely receive from many readers, taking into account the statistical uncertainty of estimating this ratio from a finite number of votes. Confidence that voting activity will remain the same is higher on a post that has received 100 votes than one that has received five.

VOTES CAST	100		50		10		5	
% upvoted	Votes cast	‘Best’ score						
100%	100-0	0.988812	50-0	0.977758	10-0	0.893898	5-0	0.799314
90%	100-10	0.858503	50-5	0.839073	10-1	0.73875		
80%	100-20	0.746583	50-10	0.722952	10-2	0.611493	5-1	0.516037
70%	100-30	0.639849	50-15	0.61395	10-3	0.497116		
60%	100-40	0.53646	50-20	0.509589	10-4	0.391742	5-2	0.299174

TABLE 9: How the number of votes cast affects the ‘best’ ordering.

In TABLE 9, as an example, of the four combinations of 80% upvoted comments (100-20, 50-10, 10-2 and 5-1), confidence is higher as the number of votes cast increases, and consequently the posts that have received more votes will be ordered above those on which fewer votes have been cast.

PERCENTAGE UPVOTED	UPVOTES-DOWNVOTES	'BEST' ORDERING SCORE
88%	44 (50-6)	0.815021
85%	85 (100-15)	0.801672
100%	5 (5-0)	0.799314
86%	43 (50-7)	0.791454
84%	84 (100-16)	0.790536

TABLE 10: *The 'best' score places some comments on which large numbers of votes have been cast, some of which are negative, above comments on which a small number of only positive votes have been cast.*

'Best' ordering is only ever applied to comments, and cannot affect the ordering of the post to which those comments are attached. Each 'top level' comment in a comment tree has a self-contained branch underneath it, and votes on its comments only affect the order of the comments in that branch.

For example:

- Post A asks a question. Three comments each offer an answer.
  - o Comment 1a gives a good answer, and is upvoted.
    - Two second-level comments, 1b and 1c, then give their support to Comment 1a. These are both also upvoted.
  - o Comment 2a, gives a different answer and attracts no votes.
  - o Comment 3a, gives a bad answer and is heavily downvoted.
    - Comment 3b corrects the bad information given in Comment 3a and gets heavily upvoted. It eventually achieves more points than comments 1b and 1c.

No matter how many upvotes Comment 3b gets, however, Comment 3a will always stay ordered below Comment 1a and 2a. ‘Best’ orders the parent comments first, and then the child comments only under their parent. Votes cast on a comment cannot change or influence the order of the post they respond to, and second level comments cannot affect the order of the first level comments.

Good answers to bad questions, or comments that correct previously incorrect ones can be ‘buried’ by the downvoting on the parent. As even a reasonably small number of early downvotes will make a post less visible to subsequent users (van Mieghem, 2011; Stoddard, 2015b), this is a disadvantage of the voting system that could be reconsidered in future platforms.

#### **4.5 Discussion**

As both a platform and a community, reddit has grown and evolved over time – a perfect demonstration of User-Centered Design. As the volume of content on it has increased, its diversification into subreddits has enabled topic-specific content to be located more easily, with the voting system indicating which information has been most favourably received. The ‘hotscore’ algorithm that is the default ordering for posts, and the ‘best’ algorithm that is the default for comments, both reward the information that is most valued by the community, creating an aggregate crowd wisdom that signposts which content is most worth visiting. The hotscore ordering also ensures a rapid turnover of Front Page items: those that are more favourably received by the community are visible for longer while ones that are less favourably received are quickly buried. Content on reddit is constantly refreshed – a valuable characteristic when one concern raised against health information on the Internet is that it can quickly become outdated.

In terms of its Collective Intelligence Genome, on the surface, reddit exists to *Create a Collection* – it is a news aggregator site that provides a single space where links to news items on other sites can be posted and shared, run largely by the *Crowd* of redditors for *Love* (in that they are not paid, and receive no other material reward or obvious *Glory* for contributing). On further analysis, however, other CI genes are at least partly expressed: though reddit is a collection of posts, the ability to comment on these posts, to discuss them and to ‘upvote’ or ‘downvote’ them depending on their perceived worth, brings in elements of *Collaboration*, with their ultimate worth to the community *Decided by Voting* (and some *Hierarchy*, as moderators can remove posts and ban posters). Posters ask questions, and reddit *Collects* answers to those questions; however, the ensuing discussion – in which answers can be debated, discussed, agreed with, disagreed with and voted on – enables the crowd of users to *Decide* collectively on which of the answers are considered to have the most merit. The individual who posted the original question is free – though may be influenced by the group discussion – to accept the group decision or to formulate their own from the same material.

The understanding of reddit set out above is relevant to this study in the following ways:

First, the ordering of information into subreddits helps to guide the information seeker to where information they want is likely to be found.

Second, if the information is not found immediately, the information seeker can make a self-post asking the precise question they want answered in an online space where people who are also interested in the same topic congregate.

Third, the ordering system used for comments – and thus for the answers received in response to a question asked in a self-post – draws on crowd wisdom to indicate which of the answers provided has the greatest value.

This has several considerations for how such a system might be used during a public health emergency. Reddit has chosen 12.5 hours (45,000 seconds) as the period for which a post's newness favours it in the ordering system over older posts; other websites could choose to prolong or shorten this. A public information site providing facts and figures on an outbreak may decide to order by 'best' or 'top' alone, foregoing the time element if information is expected to remain static and is unlikely to need updating quickly, or may give equal weight to votes no matter when they are cast. The values set by reddit are those that serve the site best: another site, with another function, would need to consider what would be best for its desired purpose.

#### **4.6 Conclusions**

This chapter has set out how reddit works and has explained its characteristics and functions. In the following chapters I will explore first what questions people asked during the West Africa Ebola PHEIC (Chapter 5) and then how reddit answered them (Chapter 6), considering as I do so the extent to which reddit is able to fulfil the information seeking needs of a population at risk from a serious infectious disease outbreak.



## **CHAPTER 5: INFORMATION REQUIREMENTS DURING A PHEIC**

### **5.1 Information requirements during a PHEIC**

During a public health emergency involving a serious infectious disease, those at risk may have many questions about how to avoid catching or spreading it, and what to do if they or their loved ones become infected (Cole and Watkins, 2015). They may try to answer these questions by seeking out information that will enable them to determine the answer, or by asking someone they think may know the answer, depending on their preferred way of learning; some people prefer to read and self-educate, while others prefer to discuss knowledge in a social, interactive environment (Biggs, 1987; Dunn et al, 1990; Fleming, 2001). In the case of health information seeking, people often employ both options, consulting a medical professional but also looking for health information online (Koch-Weser et al, 2010; Szokan, 2011). In such cases, the online information is generally used to support, rather than replace the expert opinion. As subreddits provide information and enable discussion, such forums can enable and support both learning styles.

In serious disease outbreaks, however, this journey from question to answer(s) may be subtly different. Rather than consulting a doctor only after they have become infected and/or symptoms begin to appear, individuals may start searching for information earlier, to avoid catching the disease, or simply out of curiosity about it. In October 2014, r/ebola had nearly half a million unique visitors and just under 14,000 subscribers – more subscribers than the number of cases recorded worldwide at that time, and 50 times more unique visitors.

Mathematical modelling studies suggest that awareness of a disease, passed by word-of-mouth and peer-to-peer interactions over social networks, as well as through concerted public health campaigns, may be able to slow down the

spread of a disease and help to contain local outbreaks (Funk et al, 2009; Funk and Jansen, 2013). The World Health Organization recommends proactive communication during a public health emergency that, “encourages the public to adopt protective behaviours, facilitates heightened disease surveillance, reduces confusion and allows for better allocation of resources – all of which are necessary for an effective response” (WHO 2008, p4). WHO considers this to be especially important during the early stages of an outbreak – generally in the first few days after the index case is identified, when the number of cases is low. The Organization sets out five important principles of communication during health emergencies: [1] Trust, [2] Early announcement, [3] Transparency, [4] Listening to the public’s concerns and [5] Planning (WHO, 2008, p5). However, as discussed in Chapter 1, few public health organisations have adequate plans or mechanisms for addressing point [4] – Listening to the public’s concerns – and so risk missing what these are, particularly if they are not the same as public health officials might assume.

Understanding exactly what questions were asked during the Ebola outbreak, and how health information seekers sought answers to them, demonstrates one way in which the public’s concerns can be listened to. This may help to inform not only what information is needed, but also what platform is best suited to delivering such information in a future health emergency. It will help determine what affordances are needed to maximise the perceived usefulness (Davis, 1989) of the system, which will encourage people to use it, and what might help them to trust the information accessed through it.

A health information seeker with a question or a desire to learn (e.g. What is Ebola? How do people catch it?) must decide how and/or where to seek an answer. Literature on health-seeking information suggests that while the optimal choice is to consult a doctor or medical professional (Hesse et al, 2005; Schwartz et al, 2006; Cole and Watkins, 2015), if a face-to-face meeting

with a doctor is hard to arrange, as it might well be if health sector resources are stretched or overwhelmed, the health-seeker's next question is likely to be "who else might be able to answer?".

The theories of Crowd Wisdom and Collective Intelligence both suggest the answer is out there *somewhere* and that *someone* has the information required, but also that complete information may be distributed amongst more than one person. Where a single expert is not available, the maximized choice is to ask a crowd (Aitamurto et al, 2016). The larger the crowd, the more chance there might be of a sufficiently good collective answer coming back; if no-one can provide a complete answer, combining several partial answers may be sufficient to enable a complete, or near-complete answer to be constructed – a result proposed by both the Delphi Method and Pierre Lévy's concept of distributed intelligence. If several replies are returned, the questioner may also assume that consensual answers are more likely to be correct: information from the crowd can be checked against it.

In this regard, reddit has several affordances that appear to suit the health information seeker well. First, it offers access to a potentially very large crowd that may return many complete and/or partial answers. If this assumption is correct, the next challenge for the questioner becomes how to determine which of those answers is the best, or at least better than the others: – which should be their 'maximal choice' (Sen, 1997)?

In this chapter, I will address two parts of the questioner's journey: what questions do they want to ask, and how does reddit enable them to ask these questions? I will explore whether the technology affordances of reddit do indeed provide access to a large and diverse crowd that meets the criteria of Surowiecki's clever mechanism and addresses Lévy's requirement to help users navigate the available information. I will also consider what challenges

reddit's structure and organisation might present. In the following chapter – Chapter 6 – I will focus on the quality of the information returned, and the extent to which the technology affordances of reddit help the questioner to evaluate and trust the information provided. This chapter therefore focuses on questions; the answers they receive will be analysed separately in Chapter 6.

### ***Knowledge Items + Knowledge Experts = Knowledge Communities***

An important technology affordance of reddit is that it is a *Collection* – or series of *Collections* – of candidate knowledge items (the existing posts and comments) and of candidate knowledge experts (the reddit users) (Kassing et al, 2015). Knowledge items and knowledge experts specific to certain topics congregate in subreddits dedicated to those topics – such as r/ebola – which I consider adds a third category to Kassing et al's classification: that of candidate knowledge *communities*, where knowledge items and experts are likely to be found. Kassing et al did not consider the role of topic-specific subreddits in ordering and aiding the identification of either knowledge items or knowledge experts; I build on their work by showing how this can influence health information-seeking behaviour by signposting users to where they are most likely to be able to extract the information they seek.

The structure of reddit, described in detail in Chapter 4, affords the user two options for answering their questions: they can read the existing posts to see whether these, or the comments on them, contain the answer they want (as someone else may have asked the same question previously in this way, and had that question answered), or they can ask their question directly, in a self-post, and wait for answers to be returned. A user can search for information across the whole of reddit, but if they wish to ask a question directly, they must choose a specific subreddit in which to pose it. This could be one covering a broad topic area, such as r/news or r/science, or a more focussed one such as r/ebola or r/zika.

## 5.2 Questions during the 2014-16 *Zaire ebolavirus* outbreak

The serious and widespread outbreak of *Zaire ebolavirus* that emerged in the West African countries of Guinea, Sierra Leone and Liberia (as well as smaller outbreaks in nearby Nigeria and Mali) in early 2014, and the resultant Public Health Emergency of International Concern lasting from 8<sup>th</sup> August 2014 to 29<sup>th</sup> March 2016, (WHO, 2016), provided a unique opportunity within the lifetime of this study to ask individuals in the affected regions what questions they had and where (and to whom) they looked for answers as the outbreak unfolded. The existence of a subreddit dedicated to discussing Ebola, r/ebola, has enabled me to observe the questions that were being asked in self-posts across the entire period of the outbreak and to compare questions asked at different times and in different contexts. Information collected from these Ebola witnesses provides qualitative data from a small number of people with unique experiences that can be compared to and contrasted with quantitative data from a large public platform. As the behaviour and opinions people self-report in interviews is not always entirely reliable (Alshenqeeti, 2014), the permanent record of the health-seeking behaviour of the r/ebola forum, preserved in the posts that remain on reddit, offers a valuable opportunity for comparison with the reported – and potentially more subjective – opinions of the Ebola witnesses in West Africa I interviewed.

### ***Methodology: Determining the questions people asked***

The methodology for identifying questions asked and how answers were sought and obtained has been discussed in Chapter 3, but to recap: qualitative research consisted of semi-structured interviews: 14 were conducted between 8<sup>th</sup> July and 15<sup>th</sup> November 2014 with NGO workers and employees of international companies who were stationed in, or had very recently returned from, Liberia and Sierra Leone, and a further five were conducted with

moderators of the r/ebola subreddit between 14<sup>th</sup> January and 7<sup>th</sup> April 2016. Quantitative data comes from 466 self-posts that were submitted to r/ebola between 30<sup>th</sup> July 2014 and 30<sup>th</sup> June 2016, and which either had a question mark or the word ‘question’ in their title, suggesting that the poster was directly asking a question (such as ‘Will the Ebola vaccine work?’, ‘Question about cases and deaths?’ or ‘Ebola question for knowledgeable virologist or pathologist’) or for which the first line of the text, shown below the title on the search results, indicated that the poster was asking a question. All questions were taken from self-posts, and while there are likely to be many additional questions raised in comments made in reply to these and other posts, it has not been within the resources of this study to identify and analyse these. Doing so would be a valuable future study.

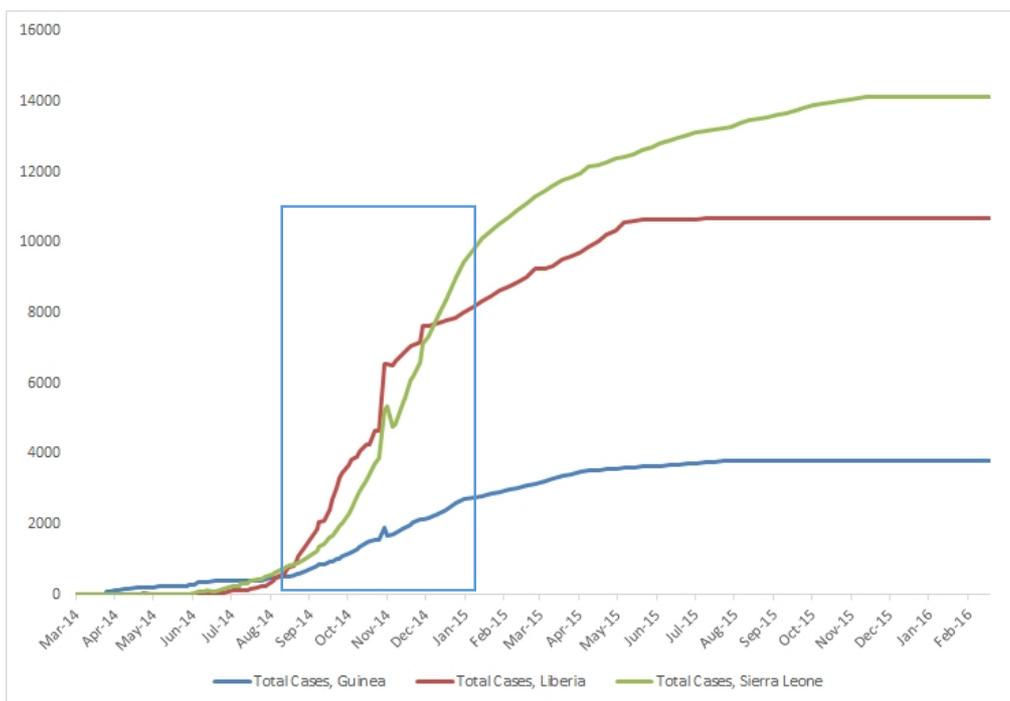


FIG 21: *Timeline of Ebola cases. The box —  — indicates the exponential growth in the number of cases in West Africa during the period of the interviews with the West Africa Group. Of the 466 questions analysed from r/ebola, 418 (94%) were asked during this same period. Image ©CDC (2015)*

The period during which the WAG interviews took place and the r/ebola questions were posted corresponds to the sudden exponential growth of the disease outbreak from July-October 2014, the crucial early stage in which WHO has identified that strong public information is vital to help contain and slow the spread of the disease. My study takes advantage of a unique opportunity to analyse the questions asked by the WAG Ebola witnesses, reddit moderators and the reddit user community during this period, and to compare them with each other to identify themes and synergies.

As described in Chapter 3, 14 interviews were conducted with people who were currently in, or had very recently returned from, West Africa during the time the outbreak was escalating. This enabled discussion of what type of questions they had had regarding Ebola and its spread.

Some of the questions identified were direct responses to an overt question asked in the interview – which can be found at Appendix I – (e.g. Question 4: “What did you most want to know?”), while others were extracted from the interviews where a question was indicated in a more general response, e.g. “I wanted to know exactly details about the case and the suspected case”, [WAG07] in response to Q3: “What were your first thoughts [when you heard there may be cases of Ebola in Sierra Leone]?” This resulted in a series of questions that the interviewees reported they had either had themselves or which they remembered friends and colleagues asking.

### **5.3 Context Shift: Far at Risk, Near at Risk, Real at Risk**

An initial finding from the interviews with the WAG was that the context in which questions were being asked was not static, a phenomenon I will label ‘Context Shift’. The interviews were cross-sectional rather than longitudinal – each interviewee was only interviewed once – but they were interviewed over

a period during which the number of cases in West Africa was increasing. As the outbreak progressed, the individuals interviewed later, and particularly four who had been in direct contact with people who were infected with Ebola, reported different behaviour to those interviewed earlier on, who had not directly experienced cases of the disease. Interviewees also spoke of how their behaviour had changed over the months since the outbreak had started.

The earlier interviewees reported limited concern that the outbreak was happening, particularly when cases were geographically and socially far away from them, while later interviewees reported more concern and caution.

*“In mid-May, something like that, I didn’t really have a whole lot of concerns. I wasn’t really thinking of moving. It was more when it came to my own area, it started to be of slightly more concern to me.”*

[WAG01, interviewed in July 2014]

*“It was in the back of our minds, constantly worry and questioning. Should we go out for dinner? Do we go out dancing anymore?”*

[WAG12, interviewed in October 2014]

*“On the street where there’s a farmer’s market, I don’t go to it anymore because of so many people on the street [...] Before I went to a church where I was very involved with Liberian friends and colleagues; I would clean the church and clean the bathrooms. I don’t do that now, I don’t have any contact any more, I shop at Western stores, not from street stalls.”* [WAG13, interviewed in October 2014]

Those interviewed later were more likely to ask questions about whether it was safe to remain in West Africa, what behaviours they should adapt if they choose to do so and what protective behaviours they should adopt to avoid

catching the disease. The questions changed from general ones about the disease to ones that could provide the advice needed to decide between different options, such as Remain in West Africa vs. Leave West Africa. This seemed to be strongly influenced by how ‘at risk’ the interviewees personally felt from the disease. At the start of the interview process, in July 2014, cases were largely confined to less developed rural areas but as the interviews progressed over the following four months, increasing numbers of cases were observed in the towns and cities where the interviewees were stationed, and those interviewed later in the study were more likely to know people who had become infected. Four of the five final interviewees had been in direct contact with an individual at a time when they were infected with Ebola. The questions they asked, and their attitude to the disease changed noticeably with this differing ‘nearness’ of the disease. I categorised three distinct stages of risk from this – Far at Risk, Near at Risk and Real at Risk, and assigned interviewees to a category, shown in TABLE 11.

WEST AFRICA GROUP			
Interview ID	Interview date	Country	Degree at risk?
WAG01	08/07/2014	Sierra Leone	Near
WAG02	09/07/2014	Liberia	Near
WAG03	11/07/2014	Sierra Leone	Near
WAG04	18/07/2014	Liberia	Near
WAG05	28/07/2014	Liberia	Near
WAG06	31/07/2014	Sierra Leone	Near
WAG07	11/08/2014	Sierra Leone	Near
WAG08	02/09/2014	Liberia	Near
WAG09	12/09/2014	Liberia	Near
WAG10	15/09/2014	Liberia	Real
WAG11	06/10/2014	Liberia	Real
WAG12	30/10/2014	Liberia	Real
WAG13	31/10/2014	Liberia	Near
WAG14	05/11/2014	Liberia	Real

TABLE 11: *Interview dates and level of risk. ‘Real at Risk’ indicates those who had been in direct contact with someone who was infected; Near at Risk indicates those living in towns or cities where Ebola cases had been recorded.*

The following table gives a broad overview of the characteristics of each category, and how this compares across the categories.

FAR AT RISK	NEAR AT RISK	REAL AT RISK
The individual is considerably geographically distanced from the outbreak with no real danger of becoming infected in the future unless there is a significant change in the outbreak trajectory.	The individual is resident in an area where some cases have been recorded, but not in specific places they usually frequent. No cases have occurred amongst friends or co-workers with whom they socialise.	Cases have been diagnosed within the individual's immediate social circle or amongst colleagues. They may be at genuine risk of contracting Ebola unless they take active steps to avoid it.
The individual is highly unlikely to be at genuine risk of contracting Ebola; cases may be present in the country but the interviewee is distanced from transmission chains.	The individual is unlikely to be at genuine risk of contracting Ebola but it is becoming prudent to consider behavioural adjustments to avoid the risk of infection.	The individual has been in direct contact with someone who was symptomatic with Ebola at the time and/or who was diagnosed shortly afterwards.
Also includes geographically distanced friends and family of those at near or real-risk who may put pressure on them to leave the region.	The interviewee has not come into direct contact with Ebola patients nor knows anyone personally who has contracted Ebola.	The interviewee needs to seriously consider how they would react if they, or family began showing symptoms of what might be Ebola.

TABLE 12: *Adapted from Table 3, Far at Risk, Near at Risk and Real at Risk, in Cole and Watkins (2015).*

All of those interviewed from West Africa had, at some point, belonged in the category 'Near at Risk', with cases of Ebola occurring in the town where they lived and/or worked, or to regions they had travelled through. The later interviewees were more likely to be in the 'Real at Risk' category, with four of the five having been close to someone who was infected and therefore in genuine danger of contracting Ebola.

Of the four who had been in the Real at Risk category: one interviewee was friendly with two U.S. aid workers who contracted Ebola, and had seen them

both during the period they were infected but asymptomatic (and therefore not contagious); one worked with the Liberian Ministry of Health and had been at a meeting with a colleague who had subsequently been diagnosed; one had travelled in a taxi with a symptomatic colleague, trying to find a hospital that would take him; and the fourth had been visited at home by a friend whose family had died of Ebola and who later succumbed himself (and would have been infected at the time of the visit).

The interviewees described their information requirements at the different stages of the outbreak they had experienced and six (WAG02, 03, 07, 08, 09 and 011) who were not Real at Risk speculated about what questions they might ask if they did move into that category, including what events might trigger this change in perceived risk status.

The interviewees were not aware of the distinction I later assigned to these stages and did not talk about their information seeking with specific regard to them; the distinction between what questions they asked when 'Far at Risk', 'Near at Risk' and 'Real at Risk' is mine alone, and therefore subjective. It is also my subjective judgement as to which incidents are described as sitting at the boundary between one of stage and another, for example a colleague or close friend being diagnosed with Ebola signalling a move from the 'Near at Risk' into the 'Real at Risk' group.

I am also aware that the community I was interviewing were being consulted during a serious disease outbreak that was already taking place (though six of the interviews were conducted before a PHEIC was officially declared). Prior to any outbreak, there will of course be a long pre-outbreak phase (for example, we may currently be in pre-outbreak phases of MERS and the next pandemic flu). There will also be post-outbreak phases that were beyond the timespan of my interview period. In these additional phases, people are likely

to have different questions and information requirements again. Though the data available does not enable analysis of these phases for the WAG, there is some option to do this with the r/ebola community and to see how the subreddit has further evolved since the outbreak has been declared over.

### ***Questions when Far at Risk***

The Far at Risk stage, during which there were recorded cases in the mainly rural areas of Sierra Leone and Liberia but not in the cities where the WAG interviewees were based, was categorised by a relatively narrow spread of questions, largely concerned with country-level statistics on the number of cases, scientific information on the nature of the virus and how it spreads, and the history of the virus and previous outbreaks, as well as a strategic interest in what the national governments were doing about it.

Ten of the interviewees (WAG 01, 02, 04, 05, 07, 08, 09, 010, 012, 014) reported looking to familiar information sources they already knew and trusted, including the World Health Organisation (WHO), the U.S. Centers for Disease Prevention and Control (CDC), national Ministry of Health websites (of the country in which they held citizenship, as well as the one where they were stationed), their own or their company's private healthcare/health insurance provider, and international media they reported to trust, such as the BBC or Al-Jazeera. As the outbreak progressed and their interest increased, ten reported that they began to seek out more information (WAG 03, 05, 07, 06, 08, 09, 010, 011, 012, 014). In general, they consulted the same sources more often, and more directly, but they also began searching out new sources of information.

Eleven (79%) of the interviewees reported that once they were aware of cases in the country, they actively searched out the latest reports on websites such as

the BBC or WHO rather than waiting for regularly broadcast news programmes or Internet news headlines. They began to check for updates on the number of cases every day. Their preferred information sources were national and international institutions – 11 mentioned WHO, five referred to CDC – suggesting that expertise and authority were important credibility markers. They were interested in facts and figures: scientific information on the nature of the disease and statistics on the number of cases.

Static information pages that set out established wisdom on the topic seem to have been perfectly adequate to answer the broad questions they had about what Ebola was, where cases were occurring, how it was spread and what symptoms those infected displayed. This can be described as information seekers actively seeking, but passively absorbing information; they did not seek to interact with the information, or require a two-way communication with its authors. They seemed happy to read and absorb.

The news aggregator characteristic of reddit, which provides the ability to post links to information from a variety of sources, could meet these needs well. Subject-specific subreddits such as r/ebola provide a portal to information from a variety of already trusted sources.

In the early stages of the outbreak, all but a handful of r/ebola posts were links to official and professionally-produced sources of information (see TABLE 13): of 103 posts between 1<sup>st</sup> and 31<sup>st</sup> July 2014, 64% linked to professional media, 16% linked to professional health agencies. Only 2% were self-posts.

The first self-post appears on the subreddit on 30<sup>th</sup> July, after which the percentage increases as the outbreak progresses: 16% of posts made on 30<sup>th</sup> August are self-posts, rising to a high of 38% of all posts made on 15<sup>th</sup> October 2014.

TYPE	No OF POSTS	ORGANISATIONS
African media	23	Accra Report (3), All Africa (12), Awoko SL (2), Front Page Africa (3), Nigeria Online (1), Premium Times (1) Sierra Leone (1)
Professional Health Organisations and National Ministries	16	CDC (1), ECDC (1), Republic of Togo (1), Sierra Leone Ministry of Health (5), WHO (8)
Professional international broadcast media	15	Al-Jazeera (2), BBC (4), CBS (1), CBC (2), CNN (2), CTV (3), Fox News (1)
International news agencies	14	AP (3), Reuters (8), NPR (3)
Professional media magazines	8	Bloomberg (2), Businessweek (1), New Vision (2), Salon (1), The Wire (1), McLeans (1)
Professional (non-African) newspapers	7	Daily Mail (1), Guardian (1), Herald Tribune (1), Hindustani Times (1), LA Times (1), New York Times (1), Times Colonist (1)
NGOs	7	MSF (3), Reliefweb (4)
Other	6	TV documentary (1), WHitv (1), Moderator post (1), Wikipedia (2), Reconomics (1)
Academic	3	Promed (2), NHAC (1)
Prof travel news	2	Eturbonews (2)
Self post	2	

TABLE 13: *Source of link posts on r/ebola in July 2014*

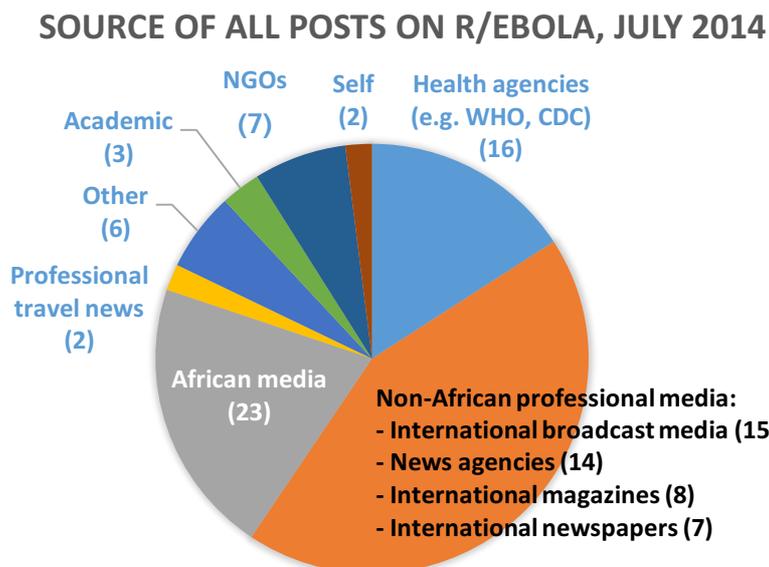


FIG 22: *Source of all posts on r/ebola in July 2014. Self-posts were rare until later in the outbreak.*

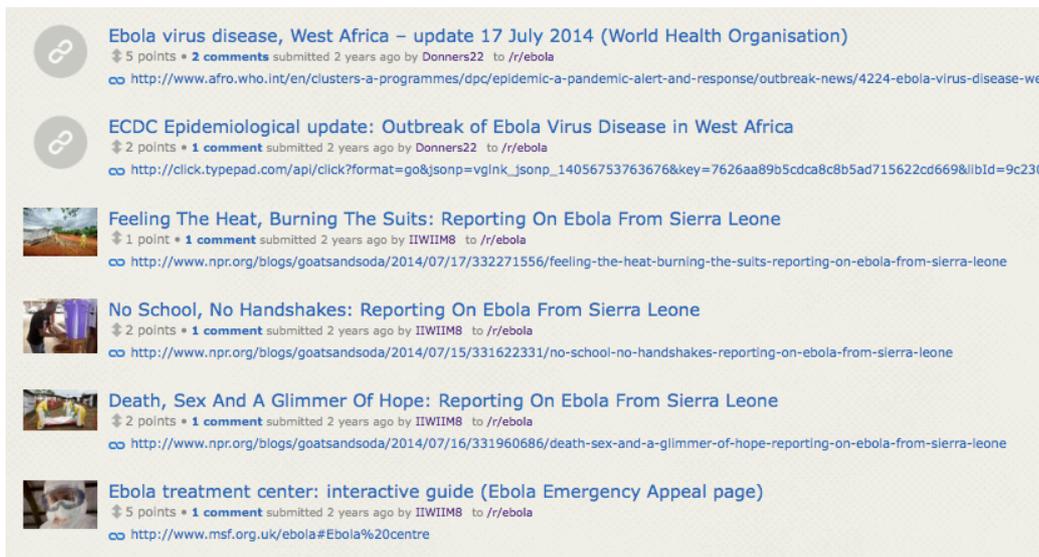


FIG 23: Link posts on r/ebola from early in the outbreak, in July 2014, to organisations including MSF, ECDC, WHO and the news syndicate NPR. This matches the type of sources reported by information seekers in the WAG group – official organisations and respected professional international media. The site’s first self-post<sup>6</sup> does not appear until 30<sup>th</sup> July.

In general, WAG interviewees did not remember the precise medium through which they had first learned of the Ebola outbreak or from where they had received the information that had pushed them towards more actively seeking additional knowledge. They received information during the early, Far at Risk phase through media channels they consulted routinely and which covered early events as part of general news or current affairs programmes. Their recall of where they had first heard that there were cases of Ebola in West Africa was, generally, no more specific than this – “BBC, Al-Jazeera, something along those lines, on the web or on TV” [WAG13]. The original source of the information – such as WHO, or a friend working in a local

<sup>6</sup> [https://www.reddit.com/r/ebola/comments/2c4v8g/are\\_there\\_any\\_culturally\\_adapted\\_sites\\_for/](https://www.reddit.com/r/ebola/comments/2c4v8g/are_there_any_culturally_adapted_sites_for/) There are technically a small number of ‘self-posts’ prior to this but these contain links to external websites, suggesting the poster was unfamiliar with the posting process and while they register as self-posts, the poster’s intention was to submit a link. The self-post on 30<sup>th</sup> July, from a user asking if anyone is aware of websites that have been culturally adapted for Liberians, as he finds the WHO website too technical, is the first time a reddit user poses a question to the r/ebola community.

hospital – was sometimes recalled but not whether they had received it through TV, a website, a newspaper, social media or word of mouth.

In terms of the four characteristics Wathen and Burkell (2002) identify as contributing to the credibility of a message – source characteristics (who originated the message); message characteristics (relevance to the context and how well it is presented); medium (the channel or system through which the message is transmitted and received); receiver characteristics (relating to the person who receives the message) – the interviewees were receiving information through mediums they already trusted. This led them to take notice of the message, and to assign credibility to its content.

Once they began actively seeking out information, the medium became more relevant, as they initially sought additional information through platforms they already knew and trusted. This suggests that an important affordance of a platform is familiarity – during the early Far at Risk stage, it may be better to push out public health messages across already popular and widely-used platforms than to introduce a new one for which trust will need to be built.

### ***Questions When Near at Risk***

As the outbreak spread and cases began to appear in major cities in West Africa, including those in which the WAG interviewees were based, such as Monrovia in Liberia and Freetown in Sierra Leone, the questions asked began to change. Nine interviewees (WAG01, 02, 04, 05, 06, 08, 09, 012, 014) specifically mentioned wanting more detail, as the disease came closer to their location, about the number and location of cases. Where exactly had cases occurred? How far from where they were based was this? How many cases were confirmed and how many were suspected? Why were some of the cases only suspected? How soon would suspected cases be confirmed?

The driver for the questions also seems to change. In the Far at Risk group, interest seems to have been down to little more than general curiosity but in the Near at Risk group there is a shift towards more practical concerns about how events might affect the interviewees directly: were there cases in areas they were due to travel to or through, for example, and if so, should they amend their travel plans? Were cases coming close enough to where they were stationed that they should consider leaving West Africa? Was their organisation – or organisations carrying out similar work – shutting down operations in other areas of the country or in other countries affected and how might this affect their employment contracts? This suggests that an important platform affordance during this phase could be the ability to discuss various options and to receive answers related to all, which can then be evaluated and compared against one another to help decide which option is ‘best’.

As well as scientific information on how the disease can spread, the interviewees reported beginning to want more detailed information on issues such as how specifically the outbreak was spreading: did the confirmed cases in the cities represent people who had become infected in rural areas and travelled to the cities, or was there local transmission in the urban area [WAG03]? How were people catching it: could it be transmitted by a waiter in a restaurant [WAG01], or by touching a doorknob [WAG012] or through sweat [WAG01; 011]? By sitting next to someone on public transport [WAG08; 010], or brushing against someone in the street [WAG013]?

They become more interested in how quickly the outbreak was spreading, and whether localised outbreaks were being contained or if they seemed to be harder to control. How was this determined, and what measures were being taken to ensure the disease could not spread further? This resulted in a desire to see more immediate situation reports from the areas (and hospitals) where cases were being recorded, including how the local population and healthcare

systems were reacting, as well as just the facts and figures. This led to an expansion of the pool of information sources consulted. There was a greater emphasis on actively seeking information from the media – particularly local media, including newspapers and radio shows – as well as from more official sources. They considered media sources to be quicker to react and to provide more immediate information. Ten interviewees [WAG01, 02, 03, 05, 06, 07, 08, 010, 011, 012] also reported sharing information they found with friends over social media, suggesting that they were adding to the collective information pool as well as just taking from it: they were becoming active participants in the Wise Crowd.

### **Context Shift Consequence I: Seeking Knowledge Experts**

A particularly important development in the Near at Risk phase is that interviewees reported that they, and their friends and colleagues, wanted to be able to ask questions (even ones that might seem stupid) not just be given facts and figures. For the first time, they were seeking people who could contribute information that had not already been deposited in the collective store.

Interviewees wanted to be able to relate this expert knowledge to the information provided by media and official sources, but nonetheless it signifies an important change in their information-seeking behaviour and a noticeable Context Shift between the information requirements of the Far-at-Risk and Near-at-Risk phases. This Context Shift brought with it several consequences for the information seeker.

The closer the (perceived) risk came, the more the WAG interviewees wanted information from people who were directly involved in dealing with the outbreak – such as doctors or nurses working in Ebola Treatment Centres – or someone personally known to them who was thought to have direct experience or expert knowledge. Two interviewees [WAG01 and WAG11] mentioned the value of information that came directly from friends working in healthcare

facilities where Ebola patients were being treated, or in the Liberian Ministry of Health. This extended to colleagues or friends who had experience from previous but similar situations. Three interviewees [WAG06, 07 and 13] referred to the value of information provided by colleagues who had worked through previous Ebola outbreaks in the Democratic Republic of Congo and Uganda, with one, [WAG06], reporting that his colleagues looked to him for advice and a kind of leadership based on his previous experience.

Two of the later interviewees [WAG012 and 014] reported the value of doctors and scientists appearing as guests on local radio phone-in radio shows, and nine [WAG02, 05, 06, 07, 08, 09, 011, 012, 013] described briefings by doctors and public health experts, organised by their employers, as valuable.

### **Context Shift consequence II: increased granularity of information**

The second consequence of the Context Shift was that the interviewees wanted more specific answers. Their information needs expanded to wanting to know what certain descriptions of symptoms meant, for example:

*“[T]he information says you have to be acutely ill for the levels of Ebola virus in your body to be high enough to transmit, but I would like to know more information on what that actually means – terms like ‘acutely’ or ‘gravely ill’. What does this mean? Could they still be walking around? Are they bedridden? I’d like it to be more precise, so I know what to avoid.” [WAG02]*

There was also concern over how such symptoms could be differentiated from other common diseases such as malaria (also seen in questions posed to CDC: Goodman et al, 2015). Interviewees were nervous of seeking medical help for any complaint in case suspicion of Ebola resulted in them being sent to Ebola treatment centres, where they would be exposed to the disease.

INFORMATION GIVEN BY OFFICIAL SOURCES	QUESTIONS ASKED /INFORMATION REQUIRED DURING THE NEAR AT RISK PHASE
Medical description of symptoms	Clearer description of medical terms: what does 'gravely ill' mean? Are those infected able to walk, or are they confined to bed? Would they be able to travel on a bus? Do they look ill? What is it like that I'm familiar with (e.g. Lassa fever)? Why/how is it different? How will I (and the doctor) know it's not [Lassa]? What if I have malaria – might it be misdiagnosed as Ebola?
Geographic location (towns and villages) where cases have occurred	How close to there is it safe to travel? Are the train/bus stations safe? Should I think about moving further away? Is it coming closer? Did the infected cases catch it there or already have it when they travelled to the area?
How many cases (numbers) have been recorded?	How many dead? How many recovered? Why are some cases 'suspected' rather than 'confirmed'? What sort of people have caught it? What will happen (to me) if I catch it?
Scientific information about how the disease is transmitted; what are its transmission characteristics?	Who is spreading it? Why did healthcare workers (who should have known how to protect themselves) catch it? What if someone sneezes on me? Can I catch it from sitting next to someone in a taxi? From a doorknob? From food? What can I do to protect myself?
Delayed but accurate figures	Earlier, estimated figures (with honesty about readjustments and corrections to earlier estimates)

TABLE 14: *Interviewees wanted information explained more clearly.*

### **Context Shift consequence III: increased size of the wise crowd**

A third consequence of Context Shift is the breadth of information sources the interviewees now chose to consult and the nature and characteristics of those sources. All 14 of the interviewees mentioned sourcing information across a variety of media and sources (including WHO, CDC, BBC, Al-Jazeera, CNN, local radio, Facebook, Twitter, the ministry of health of their homes country and the country where they were based, FCO and others). The crowd from which answers were being sought expanded dramatically. None of the interviewees reported finding this overloading or confusing but as they began to consult less trusted sources, such as social media, and hear more rumours, the desire to compare sources to see if there was consensus increased.

All of those interviewed reported consuming (and wanting to consume) information from many sources across many platforms. The types of sources can, however, be grouped into five broad categories:

(1) **Official information sources**, which includes government and international agencies such as the Liberian Ministry of Health, the American Embassy in Liberia, the Centers for Disease Prevention and Control (CDC) and the World Health Organization (WHO). Info from such sources could be received directly – e.g. from the WHO website – or indirectly, through media sources and news agencies.

(2) **Professional international media**, including the BBC World Service, CNN and Al-Jazeera, the websites of professional newspapers such as the *Washington Post* and *The Times*, and news agencies such as Reuters and Associated Press. In many cases, the media relayed information originally produced by the official information sources described above, as well as reporting directly from the affected areas.

(3) **Local professional media** including newspapers and radio shows. This largely included reporting from the local areas and included more ‘human interest’ stories of how communities were being affected and how local people were coping than could be found in the international media reports. They would also give more localised information on where cases were occurring and in what numbers.

(4) **Informal information sources**, including friends and family, from whom communication was received by word of mouth, by e-mail or over social media platforms such as Twitter and Facebook.

(5) **Formal, but personal information** sources, such as interviewees’ own doctor or medical insurance provider, or official briefings provided by employers.

Once interviewees were actively seeking out information, they demonstrated a stronger awareness of the channel characteristics of the mediums they used than had been the case in the Far at Risk stage. Which channels they preferred and valued most changed as the outbreak progressed, as did how they trusted and processed information from each. This was driven largely by whether speed or accuracy was considered more desirable for the current context: this is further evidence of Context Shift.

All the WAG group reported using more than one category of platform simultaneously and consulting more than one source: no-one reported only listening to radio broadcasts, or only reading information provided by WHO. Information heard through informal sources, and to some extent local media, was generally considered to be rumour, which interviewees would try to confirm through official sources and the more trusted professional media. This suggests both a broadening of the Crowd being consulted and, possibly, an insurance measure to guard against a single consulted ‘expert’ being wrong (Aitamurto, 2016). Previous studies on health-seeking information (Whitelaw et al, 2014; Diviani et al, 2015) have identified a desire to check likely accuracy by looking for agreement across multiple information sources: this is upheld by the reported behaviour of those I interviewed.

	OFFICIAL SOURCES	INTERNATIONAL MEDIA	PROFESSIONAL LOCAL MEDIA	INFORMAL SOURCES	PERSONAL, FORMAL
Far at risk					
Near at risk					
Real at risk					

TABLE 15: Preferred sources of information at different stages of the outbreak: Dark green = Most preferred; Light green = middle preference; White = less preference

Ten interviewees [WAG 01, 02, 04, 06, 07, 08, 09, 010, 012, 014] described deliberately seeking out multiple sources as they moved from Far at Risk to Near at Risk. The following reply is an example:

*“I was just reading anything attached to it, I was reading all of the media on it, everything I could find on it from different sources – Google, journal articles ... all the media, I was reading everything I could find.”* [WAG13]

The reasons for this are summarised in the following reply:

*“[P]art of the problem was that the government and the Ministry of Health wouldn't confirm anything until they'd got the test results back, but everyone knew when there was a suspected case. Everyone would sit around and call who they knew at the hospital and try this, that and the other to figure it out.”* [WAG01]

This did not confuse the interviewees, nor cause information overload, but helped to reassure people. As one interviewee said, “I know that I can just Google something, and it's just at my fingertips” [WAG10].

As well as turning to familiar brands such as the World Health Organization (mentioned by 11) and CDC Website (mentioned by five), which had been important in the Far at Risk stage, four interviewees [WAG02, 07, 08, 09] had also accessed known health-risk information websites such as the National Travel Health Network and Centre (NaTHNaC, [nathnac.net](http://nathnac.net)) and International SOS ([www.internationalsos.com](http://www.internationalsos.com)). This is consistent with findings by Luo and Najdawi (2004) which suggest that branding has the most significant effect on trust building online.

Ten [WAG03, 05, 06, 07, 08, 09, 010, 011, 012, 014] reported looking for information more actively and more often as the outbreak progressed, such as searching for news stories online ahead of scheduled TV and radio news programmes.

This method of distributed information seeking has comparisons with the Delphi Method, in which different experts with partial knowledge contribute to the whole; government press offices may have better information about officially reported cases while local media may have a better understanding of how health policies are being implemented and received on the ground. In the Near at Risk phase, interviewees wanted both. Comparisons can also be drawn with Surowiecki's approach to Crowd Wisdom: if enough people have some partial understanding, the aggregated best guess of many of them is likely to be close to the truth. An example here would be the number of suspected cases, or the extent of the area affected, being reported by different sources.

#### **Context Shift consequence IV: timeliness of information**

A fourth consequence of the Context Shift, and another reason reported for a preference to consult multiple sources, is a shift in the balance of accuracy vs. speed and which interviewees considered more valuable. Seven interviewees [WAG01, 03, 04, 06, 07, 010, 012] considered larger organisations (and to some extent international media) to be slow to respond. The reasons for this were generally appreciated but as the risk came closer and more immediate, people wanted information quickly. One interviewee, for example, reported:

*“...some [friends] worked in public health sectors ... health departments, and generally I would trust them over anyone else. And they also had information much, much quicker. It would take days for someone like BBC or WHO to publish anything.” [WAG01]*

Another reported:

*“The government didn’t necessarily want to share that detail of information ... through [informal] networks we found out a lot more information than was publicly available.” [WAG07]*

While interviewees described the media as tending to sensationalise, there was an acknowledgement (and acceptance) that different sources provide different types of information: the media are more likely to report quickly, but are less likely to be accurate than information from the WHO or CDC. Media organisations were generally considered more likely to be listening to people’s concerns and trying to address them, facilitating an open information exchange on those concerns as well as offering potential solutions to them. These characteristics were both seen as valuable.

INFO SOURCE	SPEED	TRUSTWORTHINESS	CONTENT
Official information (eg WHO, CDC)	Slow	Trustworthy; confirmed; accurate	Facts, figures, statistics, big picture
Professional media (international)	Reasonably quick	Largely trustworthy and accurate (but needs to be confirmed); honest; can be alarmist but largely responsible	Contextual, emotional. International coverage shows issues are ‘real’
Professional media (local)	Reasonably quick	Often speculative and may be sensationalist; may be less responsible than international media; needs to be confirmed	Emotional, more personal to the local context. Better Understanding of local challenges
Informal information	Immediate	‘From the frontline’ but also rumour; may be highly subjective	What is really happening; personal experience
Formal information – personal	Immediate and accurate, relevant to them	Trustworthy and context aware	What you should do; advice and help in making decisions

TABLE 16: *Information sources and perceptions about them (adapted from Cole and Watkins, 2015)*

The result of this change in information seeking behaviour – to a widening of the sources consulted to include those that were known to be less reliable and less accurate, and to include new sources that may not yet have gained users’ trust – was that consistency across the multiple sources became important.

Consistency provided reassurance and helped the information to be trusted. The interviewees described putting together information from multiple sources (and multiple categories of sources) and then analysing it to construct their version of the truth. As one said: “some of it you take in, some of it you discard, some of it you take with a grain of salt, some of it was actually quite useful” [WAG04]. Another commented, “technical information that you’re getting from WHO and MSF [...] is very useful but it doesn’t get into [...] the cultural issues that surround the disease”, [WAG03] whereas the media did. In doing this, interviewees would pass on information from the sources they considered to be the most factually accurate to friends and colleagues, sharing the components of their truth construct, if not the construct itself.

Particularly reassuring was the ability to turn to consistently accurate and reliable trusted sources such as WHO, CDC and the local Health Ministry – the preferred sources from the Far at Risk phase – to check, and hopefully confirm, the information constructed from the local media sources and informal sources, even if this official information came later.

Interestingly, the interviewees did not tend to draw a distinction between face-to-face informal sources, such as talking with friends and colleagues, and communications over social media. Internal company communications, such as an organisation’s own website or internal e-mail group, was often described in the same way as informal sources by the company’s own workers, but was seen more as official communication from a trusted brand by non-employees.

### ***Questions when Real at Risk***

The increased granularity observed in the WAG group who were Near at Risk continued into the WAG group who genuinely became Real at Risk. Concerns became increasingly specific to their own situation as the risk came closer: questions were no longer theoretical but more focused on what was in fact happening around them and what they should do about it.

*“[My husband] looked at me in horror and he said, “you’d take this to America?”, and I said yeah, yes, so that’s my answer [...] I guess... if I’m going to die of it, I’ll die of it, but I wouldn’t want to do it on the other side of the world, away from my kids.” [WAG011]*

*“We did barricade ourselves in. We got food... we thought we’ll do it for a week and see, and then after, we decided, this is crazy... we can either go, or stay. If we stay, we stay whole-heartedly, and get all the information we need to make sure we’re safe.” [WAG014]*

The WAG interviewees began to wonder what would happen if there was a case in their house: how would they decide who would care for an infected relative so that not all the family were put at risk of infection? Should they attempt to fly back to the U.S. before they began to show symptoms, as healthcare provision was likely to be better there? Would they be putting others at risk? Their questions became more directed towards helping them to make decisions, rather than just understanding the situation better.

Information requirements became more personal – not only where Ebola Treatment Units were located, but also what conditions were like inside them. What happened to people who had been diagnosed? How were friends and

colleagues who were choosing to leave or stay making their decisions? Would their wages be paid if they left when company policy was not to pull out?

Questions also became very specific to what could be seen at the local level, for example:

*“I never really thought about pets but a lot of people asked about pets and then I realised that Liberians were afraid of dogs... so there were sometimes bodies on the street and these dogs, they were having all these thoughts that if the dog touches the body or eats the body and then it comes and touches you can it...?”* [WAG13]

In this phase, expert knowledge that could answer scientific and medical questions became less important than the experience of people who would understand the context and be able to advise the questioner on their specific situation. Five WAG group members [WAG02, 07, 011, 012, 013] reported constricting their information circles – sharing information amongst their known friends and family, and exhibiting information more like a typical health community than disparate information seekers. They no longer wanted to throw questions out to as wide a crowd as possible – they wanted answers from ‘people like me’, showing similarity with the homophily Wang et al (2008) report being important to peer-to-peer discussion forums.

Official and international media sources were considered, “[s]terile... I mean, they were just numbers ... but what [our friends] were telling us ... it was factual, not suppressed or manipulated in anyway”. [WAG11]

This represented a further Context Shift, at the boundary between Near at Risk and Far at Risk, to that observed at the boundary between Far at Risk and Near at Risk.

## 5.4 Context Shift on r/ebola

The information seeking described above relates to a small group of international workers stationed in West Africa during the outbreak. While it is important not to assume that their experiences can be generalised to the population of the UK, or any other nation during a future disease outbreak, it has been useful to compare their reported information seeking to that observed on r/ebola and described by reddit moderators, to see what comparisons might be drawn, and how or if Context Shift affects an online community, particularly as, while the WAG interviewees represent a geographic community, r/ebola is a community of interest – a knowledge community which may hold, order and disseminate information.

From 1<sup>st</sup> June 2014 to 31<sup>st</sup> October 2014, activity on r/ebola increased considerably. During June 2014, the subreddit had only 82 unique visitors and 180 page views – an average of just over two views per unique visitor. Both the numbers of unique visitors and the average number of page views per unique visitor increased over the next few months as the outbreak became more severe: more people were coming to r/ebola (indicating they were actively seeking out new sources of information on Ebola) and, once they had found it, they consulted it more often as the outbreak progressed. This is consistent with the behaviour reported by the WAG interviewees: the users of reddit seem to have made the same Context Shift, from being passive receivers of information to active information seekers as the Ebola outbreak grew.

By October 2014, the average unique visitor referred to r/ebola more than 11 times per month, an increase from just over two visits per month in June (in contrast, the average unique visitor refers to r/diabetes five times per month<sup>7</sup>).

---

<sup>7</sup> Data supplied by r/diabetes moderators in private correspondence

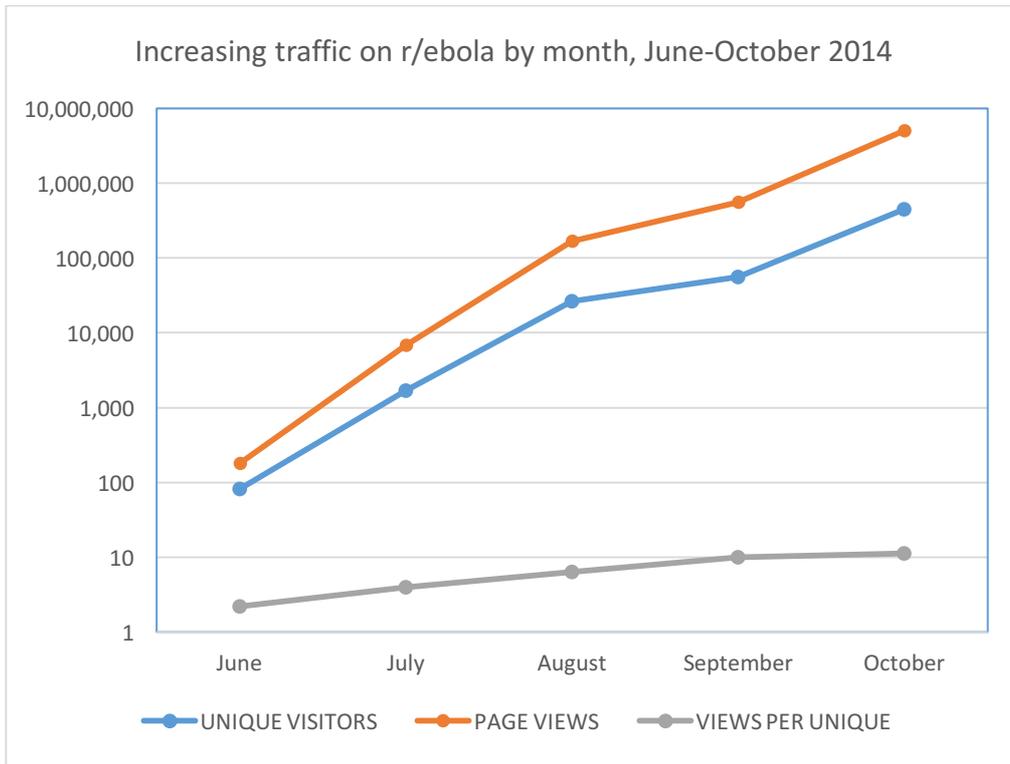


FIG 24: *Each unique visitor viewed r/ebola twice on average in June 2014, but this average increased to 11 site visits per unique visitor by October.*

	UNIQUE VISITORS	PAGE VIEWS	VIEWS PER UNIQUE
June total	82	180	2.2
July total	1,698	6,839	4.0
August total	26,195	26,195	6.4
September total	55,793	556,473	10.0
October total	446,497	5,017,046	11.2

TABLE 17: *The number of unique visitors, page views and views per unique visitor to r/ebola increased between June and October 2014.*

The permanent record represented by r/ebola has also enabled me to analyse the questions asked by the reddit community during the period 1<sup>st</sup> June 2014-30<sup>th</sup> June 2016 and compare them with those asked by the WAG group.

At the height of the outbreak, r/ebola had just under 15,000 subscribers and just under half a million unique visitors per month. By choosing r/ebola, rather than one of the much larger subreddits such as r/news (3.8 million subscribers at that time) or r/science (5.8 million), where Ebola information could also be found, the information seeker was choosing specificity over quantity. Everyone on r/ebola was expressing a conscious interest in Ebola: the 15,000 subscribers represented a potential pool of 15,000 people who might be able to answer a question themselves or point a questioner towards information or other websites that may be able to help. Not all the subscribers to r/news or r/science in the same month were specifically interested in, or potentially knowledgeable about, the Ebola outbreak. This data supports the behaviour reported by the WAG that two consequences of Context Shift were to actively seek information, and to do so across new mediums and platforms.

The existence of subreddits is a particularly useful affordance for those looking for expertise in a specific subject area. Like the WAG interviewees, the reddit users turned to an information source they were already familiar with (reddit) but started to actively look for information, rather than passively absorb it. The ordering of information into subreddits helped to signpost where that information, and people who held additional information on Ebola, was more likely to be found.

### ***Changes in size and characteristics of r/ebola***

The size of the r/ebola community grew as more reddit users sought out Ebola information. As the number of users grew, the data also suggests that the nature of the community changed. An increased number of subscribers and page views will inevitably bring an increased number of posts but the ratio of self- to link-posts also changed, suggesting that the nature of the community was changing. People became more likely to ask questions.

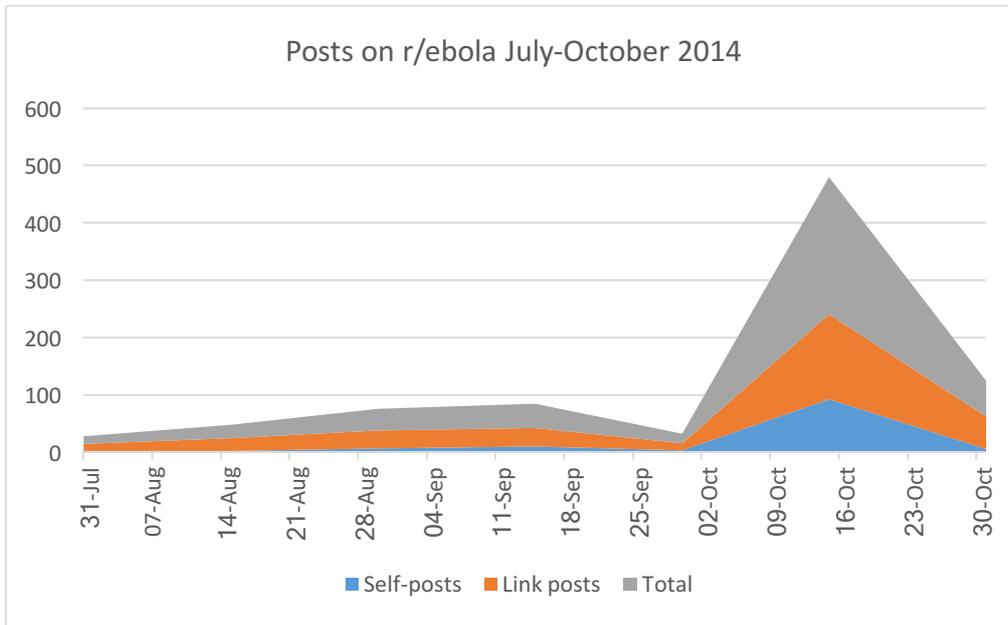


FIG 25: Total number of link-posts, self-posts and new subscribers on r/ebola. The percentage of self-posts increases from 8% in mid-August to 38% on the subreddit's busiest day (15<sup>th</sup> October 2014) but drops back to 9% by the end of October, when the outbreak is starting to be brought under control.

	TOTAL SUBSCRIBERS	SELF POSTS	LINK POSTS	TOTAL	% SELF
31 July	130	2	12	14	14
15 Aug	657	2	22	24	8
30 Aug	1049	6	32	38	16
15 Sept	1814	10	32	42	24
30 Sept	2671	3	13	16	19
15 Oct	10708	92	148	240	38
16 Oct	12003	85	198	283	30
17 Oct	12617	55	160	215	26
31 Oct	14176	5	57	63	8

TABLE 18: The percentage of self-posts on r/ebola was highest at the height of the outbreak, when the forum was also attracting many new subscribers.

The questions on r/ebola were identified from self-posts which either end in a question mark or were clearly intended to be a question. For the period before 27<sup>th</sup> August, this was determined by reading through all posts and identifying

those showing the self-post icon, but on 27<sup>th</sup> August, ‘flair’ was introduced on the subreddit, enabling posts to be filtered by the ‘self’ flair which is more likely to indicate a self-post in which a question might be asked. From then on, only posts returned after filtering for self-posts are included in the results.



FIG 26: *Icons distinguish between link posts and self-posts. In addition, self-posts may also be given a ‘self’ flair to distinguish them from posts that do not link to external content but which are scheduled interviews or discussions with guests, such as ‘Ask Me Anything’ sessions, hosted entirely on the site.*

Using these criteria, the first question appeared on r/ebola on 30<sup>th</sup> July 2014, after which the average number of questions asked per day increased dramatically from an average of 1-2 per day in August and September 2014 (38 in August and 46 in September) to between 10 and 20 a day in October 2014. Of the 466 self-post questions asked on r/ebola in the five years from its inception to 1<sup>st</sup> December 2016, 109 were asked over three days in mid-October 2014 (23.4% of all questions ever, in a period representing >0.2% of the forum’s history), and 80% were asked between 1<sup>st</sup> September and 31<sup>st</sup> October 2014 (>4% of the forum’s history).

This change in information seeking raises questions that can best be explained by examining what was driving the increasing interest in Ebola. Between July and October 2014, 11 Ebola patients were either treated or diagnosed on U.S. soil (seven were medical evacuations who were infected and became symptomatic in Africa but were flown to the U.S. for treatment; two contracted the virus in Africa and travelled to the U.S. before being diagnosed;

and one of those infected two of the healthcare workers who treated him). As these events unfolded, there were two distinct periods of rapid growth on r/ebola. The first upward curve corresponds with the first cases of Ebola on U.S. soil – aid workers infected in Africa who were flown home for treatment. The second, more vertical upward curve, corresponds to the first case diagnosed on U.S. soil – a Liberian who had travelled to the U.S. to visit family before falling ill, and who subsequently infected two healthcare workers while being treated at a U.S. hospital. These two trigger events appear to bring Ebola ‘closer’ to the largely U.S.-based reddit community, moving it from Far-at-Risk to perceived Near-at-Risk.

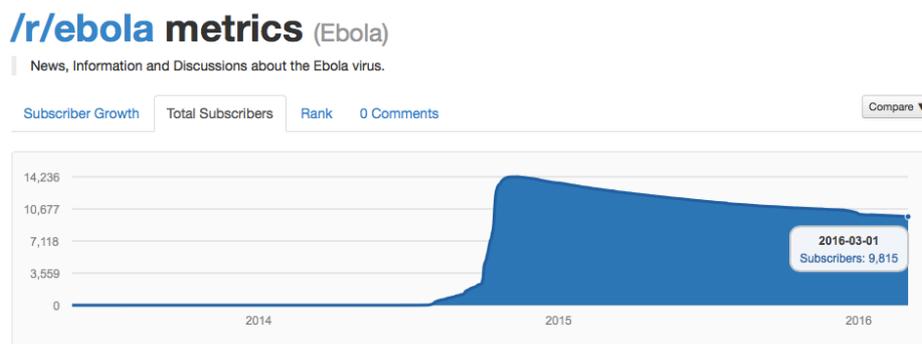


FIG 27: Timeline for the r/ebola subreddit showing subscriber growth. By July 2016, with the outbreak all but over, many subscribers have left the site.

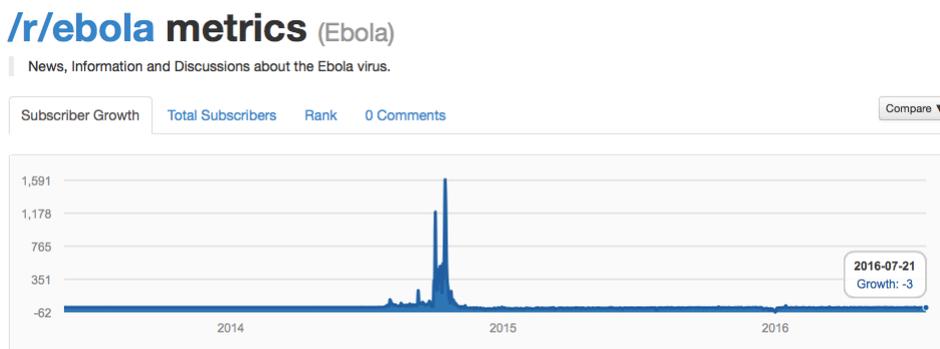


FIG 28: There are two distinct spikes in subscriber growth. The first corresponds with infected aid workers being flown back to the U.S. for treatment, the second with the first cases of Ebola diagnosed on U.S. soil.

## **5.5 Community Shift on r/ebola**

As cases moved from West Africa to U.S. soil, the largely U.S.-based reddit user community seemed to feel they may be moving from Far at Risk to Near at Risk. Not only did an increasing number of people interact with the r/ebola subreddit, and begin to ask more questions on average than community members had done previously, but the nature of the questions also changed. This shifted the nature of r/ebola from a knowledge community – in which members held and shared existing knowledge – to a community in which many members were seeking information they did not already have. I will call this phenomenon Community Shift, and explore the consequences it had for r/ebola.

### **Community Shift Consequence I: The type of questions asked**

Once the r/ebola posters start asking questions in larger numbers, the types of questions they ask bear distinct similarities to those asked by the WAG interviewees in the Far at Risk and Near at Risk phases and, as the outbreak continues, there are similarities with the type of questions observed in the WAG Real at Risk phase. It is important to realise that while approximately half of all reddit users are from the U.S., people posting on r/ebola might reside anywhere in the world and may not all have been experiencing the same phase of the outbreak at the same time. Three posters, all posting on the same day, may represent the Far, Near and Real at Risk groups simultaneously.

It is therefore more difficult to sort the questionnaires into such categories as it was for the more geographically contained WAG group, particularly as there is no detailed demographic data available for the subreddit. This is a limitation of the study, but nonetheless there was a marked change in the characteristics of the questions as the outbreak progressed.

Between July and mid-August 2014, only a handful of questions were asked on r/ebola. These tended towards more distanced, strategic questions reported by the WAG interviewees during the Far at Risk phase, on the history of the disease and scientific information on how it spreads. The paucity of questions during this period suggests that the increasing number of information seekers who were coming to r/ebola were happy with the non-interactive information provided in link posts.

Some strategic trends can be observed on reddit in the Far at Risk period that were not reported by the WAG group, such as questions about whether the media was taking the situation seriously enough (and whether people in the U.S. and other developed countries were too complacent in general), what might be likely to happen in the future (based on posters' own speculation and discussion of outbreak projections from WHO and CDC), and also several questions about how reddit users might help, through donating money, time or specific skills to the medical and relief efforts. There are questions relating to vaccines, whether survivors are immune, how this outbreak compared with other previous outbreaks of Ebola and SARS, and questions relating to whether governments, the WHO and CDC (and the information they put out) can be trusted, generally due to assumptions of incompetence/being overwhelmed rather than concerns over wilful misinformation.

As the outbreak progressed into August and September 2014, however, more questions appeared relating to the precise location and number of cases, what international governments and international agencies were doing about it, what symptoms were like and how to prevent the disease from spreading. This mirrored the type of questions asked by the WAG interviewees in the Near-at-Risk phase of the outbreak. As with the questions asked by the WAG, questions in this phase suggest a need for a diversity of expert opinions beyond that of just scientists and doctors: the questions asked related to policy,

legal positions, border controls, history and human behaviour, requiring a broad range of information from different fields of knowledge. The official information sources such as WHO and CDC preferred in the Far at Risk stage rarely cover all these areas themselves, while official information sources that did discuss additional issues such as quarantine, such as *National Geographic* (Weintraub, 2014), may not have been the ones information seekers would think to consult as they are not health-specific. The aggregation characteristic of reddit enabled information from non-health specific sources, as well as the obvious sources of Ebola information, to be linked to an Ebola-specific forum.

r/ebola also enabled those unsure of where to find an answer to ask their question and wait for responses from the potentially expert crowd (or rather, potential experts within the crowd). In the perceived Near-at-Risk phase, the community shifted from one whose members primarily contributed to the collective store of knowledge to a larger percentage who came to take from it.

As with the WAG group, there were questions relating to specific cases, including that of Patrick Sawyer, an infected Liberian-American who travelled to Nigeria and caused a small outbreak in late August 2014, but most notably about the cases that hit U.S. soil. The diagnosis of Thomas Eric Duncan in Dallas on 30<sup>th</sup> September 2014 and, by mid-October, the two onward transmissions from Duncan to healthcare workers who had treated him, had a noticeable impact. Around 10% of the questions asked on r/ebola during October 2014 were either asking for situation reports/updates on these specific cases, or for more information about what had happened/could happen in that specific case. Prior to the cases on U.S. soil, there had been some questions relating to specific cases – such as how Patrick Sawyer had been able to travel by air to Nigeria – but most had been more general, about how the disease spreads and how many cases were predicted. Typical of this change are the following questions, all asked on 15<sup>th</sup> October 2014, one of the subreddit's

busiest days, when I identified 43 individual self-post questions, the site received 370,520 page views (49,378 uniques) and gained 1,600 new subscribers, taking it over the 10,000 subscribers mark for the first time:

*“I person comes in with Ebola, 2 people get it [...] Was this a mistake or do you think this will happen a lot?”* u/TatM, 15<sup>th</sup> October 2014

*“How did the nurses get Ebola and [Thomas Eric] Duncan’s family members didn’t? Shouldn’t they have a higher probability of getting infected since they shared the same toilets and stuff?”* u/imperialdoor, 15<sup>th</sup> October 2014

*“What are the risks for plumbers serving the Texas Health Presbyterian Hospital? Can they contract Ebola from working with the waste systems?”* u/PerhapsTooHonest, 15<sup>th</sup> October 2014

*“Why wasn’t [Thomas Eric] Duncan sent to a [BioSecurity] Level 4 facility?”* u/Kytnurl, 15<sup>th</sup> October, 2014

Another significant ‘new’ type of question that appears, and suggests that members of the community are beginning to see themselves as being Near(er) at Risk, is scientific or medical questions relating to specialist topics relevant to the cases of interest, asking where to find more information, and/or what those terms mean, as well as questions about government’s role in limiting the mobility of those known or suspected to be infected, typified by the following:

*“Could you clarify what ‘monitoring’ means in terms of those in Dallas who treated the Ebola patient being monitored? CNN said 70+ people are being monitored? Are they being quarantined?”*  
[u/Spikekuji, 15<sup>th</sup> October 2014]

*“In light of the [...] decision to isolate rather than euthanize the dog of the infected nurse, I was curious what we knew about transmission of Ebola from dogs to humans?”* [u/whimbrel, 15<sup>th</sup> October 2014]

Some posters also display increasing concerns over what steps international governments are or should be taking:

*“So, for weeks [...] there has been a call for a ban on air travel for individuals from the Ebola infested areas in West Africa. However, the repeated message has been that it would have a negative impact on the efforts to contain the virus [...] So what’s the deal here, which is it? Ya (sic) can’t have it both ways.”* u/SurfaceBeneath, 15<sup>th</sup> October 2014

*“In the U.S., what rights can the CDC/Government take away from you if you have been potentially exposed? What rights can the Government take away if you have been in contact with an infected person?”* u/Nuhvok, 15<sup>th</sup> October 2014

Many of the questions express a desire to protect oneself from the risk of catching the disease, even if this means taking a harsh stance on the rights of others to, for instance, undertake international air travel or leave their home. This marks a distinct community shift from passive absorbers of information to interactive contributors in a debate that could become heated. As one r/ebola moderator [RM010] described it: “A subreddit that had initially consisted of people who had a strong interest in the subject became overwhelmingly populated by people who had a strong fear of the subject [...] One thing that we picked out of the Ebola epidemic was that often people want to have very specific questions answered – they want to have an interaction, rather than just a static Q&A, and in the absence of good access to a doctor to ask, an online discussion forum may be the best option they have”.

These conclusions were very like those I had listed in the paper on the WAG group published, prior to this interview, in the *Journal of Business and Emergency Planning* (Cole and Watkins, 2015). The affordance reddit gave of being able to ask questions in what had originally been, and largely continued to be, a news aggregator site (even at r/ebola's height, questions accounted for less than a quarter of all posts) appears to have been particularly valuable. An established information source, which could be easily identified by potential users as a likely source of knowledge experts as well as knowledge items, shifted from being a community of experts to a community with more mixed experience and expertise in which newcomers looked to the established community members for information and support – more like the communities found on chronic condition subreddits.

It is particularly interesting that some of the very specific questions unlikely to be answered by an FAQ or official sources were mentioned by both groups. Common to both groups were concerns over the transmission risk posed by stray dogs, observed on the streets of Liberia and Dallas. Such animals were perhaps common to the low socio-economic areas where Ebola was emerging but somewhat distanced from the 'ivory towers' of the scientists and health officials drafting the FAQs – Beck (1973) has shown, from whole city surveys conducted in the United States, that free-ranging dog populations are correlated with high-density, low income areas. Similarly, common to both groups was a desire to have a better understanding of how easy (or not) it would be to catch Ebola from travelling on a bus in hot, cramped conditions.

As some community members started to perceive themselves as becoming in danger of being Real-at-Risk, there was an upswing in the number of questions that related to how one should prepare for an outbreak in the local area, mainly regarding what items should be stockpiled so that going outside could be avoided:

*“Shouldn’t we be prepared? Just in case? The question is, what items do we need to be prepared?”* [u/raydar4567, 15<sup>th</sup> October 2014]

*“How can one self-isolate for 21 days? Who does the shopping? How could anyone really comply without sneaking out?”* [u/kriesedpj, 15<sup>th</sup> October 2014]

This Community Shift and its consequences were not necessarily embraced by the original r/ebola community, which had largely migrated from r/news and consisted of people interested in scientific and fact-based Ebola information.

### **Community Shift Consequence II: The impact on r/ebola moderators**

The interviews conducted with five r/ebola moderators provided valuable data not only on the types of questions being asked on the forum, but also how the platform was able to respond (or not) to the Context Shifts and Community Shifts taking place. This will be further explored in Chapter 8, but issues specific to understanding how questions were asked are discussed below.

The r/ebola moderators interviewed belonged, throughout the outbreak, exclusively to the Far at Risk category. Three were U.S. citizens, one was Australian and one was from the UK. None were based in West Africa during the outbreak, nor in the immediate vicinity of any of the U.S. or U.K. cases. Most described themselves as being interested in “fact-based discussion [...] reputable news and balanced discussion” [RM16]. The interests they described were consistent with those described by the West Africa Group in the Far at Risk stage and observed in the questions asked by the reddit community they served in the earlier months of the outbreak, consisting of facts and figures that had been verified by international organisations and professional scientists. As one of the moderators put it [RM15], “The outbreak was primarily located in West Africa, which meant that there was minimal

emotional attachment for most posters. That allowed for much more balanced and rational discussion. It was the extremely limited and brief American outbreak which caused r/ebola to go out of control”. The term ‘out of control’ refers to the rapid increase in traffic – page views and posts – to r/ebola during September and October 2014, when 11 Ebola patients were either treated or diagnosed on U.S. soil, triggering a Community Shift on the subreddit which the moderators recognized and tried (largely unsuccessfully) to resist.

At least two Community Shifts happened during the Ebola outbreak – in the first, reddit users migrated from general communities such as r/worldnews, in which Ebola was discussed amongst other topics, to the dedicated community of r/ebola. In the second, the new, larger r/ebola community shifted away from sharing factual, scientific information as a significant part of it began to speculate on media stories and to ask questions. This second shift could have been resisted or blocked, enabling r/ebola to retain its fact-based approach, or a third shift could have occurred, in which a new community might have been created where a different type of Ebola discussion could take place. Instead, r/ebola experienced a Community Shift that saw its characteristics change. The moderators, who considered the community to be Far at Risk, struggled to adapt to a significant percentage of users behaving as if they were Near at Risk. This will be explored further in Chapters 7 and 8.

REDDIT MODERATOR GROUP			
Interview date	Interview ID	Country	At risk?
14/01/2016	RM07	Australia	Far
05/02/2016	RM008	US	Far
05/02/2016	RM009	US	Far
Dec 2015-Jan 2016	RM010	UK	Far
07/04/2016	RM015	US	Far

TABLE 19: *Reddit moderators of r/ebola interviewed during this study.*

The r/ebola community never became Real at Risk, so whether the subreddit would have undergone a further Community Shift – in which the group again shifted the information sources consulted in the Near-at-Risk stage, to seek information amongst close friends and family with personal experience – can only be speculated. The behaviour reported by the WAG group suggests, however, that a reddit affordance which might prove useful in this context is the ability to start new subreddits that could be specific to defined geographic locations, such as r/ebola\_NewYork or r/ebola\_London and provide a forum for affected communities. The ability of reddit to enable such communities to be set up quickly, supported by the more experienced users of the existing r/ebola subreddit, may be a particularly valuable feature of this type of platform, especially if outbreaks happened in different regions at different times, so that one community that had gained experience of living with the disease was able to provide experiential knowledge to those still in a Near at Risk phase or just entering Real at Risk with low numbers of cases. Such a community might be able to action the information flow identified by Funk et al (2009, 2010) as being able to slow the spread of a potential outbreak.

## **5.6 Post-Risk: Context and Community Shift**

The r/ebola subreddit returned to ‘normal’ at the end of October 2014, when the speculated outbreak resulting from Thomas Eric Duncan’s onward transmissions, and the second case of a diagnosis on U.S. soil (an MSF doctor who had flown back to the U.S. before symptoms had developed) failed to materialise. Just 18 self-posts displaying questions appeared during the first two weeks of November 2014, dropping to barely two per week in the second half of November. Since then, only a dozen questions have been asked, most of which relate to whether the outbreak can be really considered to be over.

The number of subscribers has dropped slowly since then and stood at 8,585 on 1<sup>st</sup> June 2017. In December 2016, there were only 713 unique visitors and 1,555 page views (2.18 page views per unique visitor, in line with the rate in June 2014, when the forum first became active), compared with nearly half a million in October 2014. The Context Shift to Post-at-Risk has again triggered a Community Shift, or rather a return to a community more characteristic of that before the outbreak with a small number of active users, each of whom refers to the site less frequently, and makes predominantly link posts.

## **5.7 Discussion**

The interviews I conducted and forums I observed offer a better and more nuanced understanding of what questions people ask during a serious disease outbreak, allowing us to consider how a health information platform (or platforms) may be best configured to answering them. This in turn indicates which technology affordances might provide access to a wise group, potentially enabling Collective Intelligence to emerge and thus maximising the choice(s) of those using the platform.

The West Africa Ebola outbreak has provided a unique opportunity to study real questions people asked during an actual outbreak of serious infectious disease. A unique contribution to knowledge is that information requirements may not be static but rather appear to change at different phases of the outbreak. This has not previously been identified or explored in the academic literature. My research has identified three distinct phases – Far at Risk, Near at Risk and Real at Risk. The WAG group interviewees reported sourcing information differently as the outbreak progressed and the risk of contracting Ebola became more real. On the r/ebola forum, similar changes were observed as cases of the disease reached U.S. soil.

### ***Far, Near or Real: Contextualising risk***

The way in which the interviewees from the West Africa Group and reddit described and contextualised risk was broadly down to how immediate to themselves they felt it to be. Interviewees in both groups generally described their perception of risk moving from one level to another as being triggered by cases ‘leaping’ out of one area and into another, which they perceived to be closer to themselves. In the West Africa Group, this was from rural areas into urban environments – particularly the city in which they were stationed – while for the reddit community, this corresponded to cases being diagnosed on U.S. soil. The case of Thomas Eric Duncan, who was able to fly from Liberia to the U.S. and enter the country with his Ebola undetected, was turned away from hospital on his first visit, and after being admitted subsequently infected two healthcare workers, one of whom made an internal flight within the U.S. before she herself was diagnosed, drove more interest in r/ebola (measured by increased page views, self-posts and subscribers) than the repatriation of infected America healthcare workers, which may have been seen as a more controlled act in which the disease was safely contained and less likely to result in an outbreak within the U.S.

As the individuals within the WAG group experienced Context Shift, their information requirements and associated health-seeking behaviour changed. The event was fast-moving, and as people attempted to integrate a wider range of information sources that were individually less reliable and possibly not independent of one another, they sought information across a broader range of platforms which could be compared and aggregated. Neither the West Africa Group nor the reddit community minded receiving incomplete or estimated information, or information that was later corrected. Interviewees from both groups appreciated information may not always be perfect, but would rather have whatever incomplete information was available than nothing.

This Context Shift is also observed in r/ebola, but from a somewhat different direction. If we consider the subreddit to be a community comprised of the individual users, as well as a collection of posts, the community changed as the outbreak developed. It expanded – from a handful of users to around half a million – and the individuals it attracted changed from those interested in posting scientific papers and sharing fact-based information to those more interested in media speculation and asking questions about local cases, resulting in Community Shift from a small community of well-informed knowledge holders, to a larger community with a higher percentage (from 8% in July to as high as 38% in the middle of August 2014) of poorly-informed knowledge seekers wanting answers to specific questions. When Community Shift happens, or may be about to happen, the platform serving the community can adapt and embrace the Community Shift that comes with it, or reject it and encourage individuals to seek out a more appropriate platform for their new context. I explore how this process played out across reddit in Chapter 8.

Context Shift and Community Shift both drove the need for different technology affordances at the Far at Risk, Real and Risk and Near at Risk phases: a platform for static information exchange, providing facts and figures in Far at Risk; an interactive forum enabling questions to be posed to a large community in Near at Risk; and a forum serving a smaller, community support group in Real at Risk, in which experiential as well as scientific expertise is valued. These different needs could be provided by three separate platforms, each serving a community experiencing a different phase, or a single platform capable of rapid adaptation as the context of the community it serves changes.

### ***Triggers for Context Shift and Community Shift***

The observation of r/ebola suggests that individuals and communities may act on the perceived risk, rather than the actual risk, of catching a disease. It is

therefore particularly important to understand what might trigger a perceived Context Shift from one risk phase to another. This may be the period during which the opportunity to influence community behaviour and slow the spread of the disease is greatest, particularly if the questions the community will ask can be anticipated, with suitable answers – and answer providers – prepared.

Within the West Africa Group the context shift from Far at Risk to Near at Risk was triggered by cases coming geographically closer, into the city where the interviewees were based, and where they may be likely to encounter infected people on public transport, at church or in the market. The context shift from Near at Risk to Real at Risk was generally triggered by hearing that someone the interviewee knew personally, or felt an affinity to (such as an employee of the same company, even in a different regional office), had contracted the virus, as well as actual contact with Ebola sufferers.

There was a definite sense that if ‘someone like me’ can be affected, ‘I’ might be too. In this regard, it is important to know who the community sees as ‘people like me’. One WAG interviewee [WAG08] remarked: “After the two U.S. health workers got infected, that’s when everything changed – people started quitting their jobs and moving overseas, NGOs were pulling out”. Other Westerners were considered much more ‘like me’ than geographically closer African colleagues. The WAG interviewees generally did not seem overly concerned when the disease was in their country. Some began to show more concern when it was in their town but real concern came when someone they knew personally had it: “If they made a mistake, might I?” [WAG013].

Behavioural changes, such as changing shopping habits or avoiding crowded places and social gatherings, seemed to kick in at the border between the Near- and Real-at-Risk rather than at the border between Far- and Near-at-Risk, though the Far-at-Risk reddit community certainly speculated about

what they would and should do if their situation worsened. Those in the WAG group showed more concern for people they knew in higher risk categories than for themselves.

The geographically and socially closer that cases came, however, the more likely people were to start asking questions such as: Should we go out for dinner? Do we go out dancing anymore? What should we stockpile? The last three WAG interviewees [012, 013, 014] all reported that they were more cautious about going out; avoiding shaking hands and taking shared taxis was also mentioned. The last interviewee was working from home, and was barely leaving the house. Some of the concerns voiced here are shown in TABLE 20.

FAR AT RISK	NEAR AT RISK	REAL AT RISK
Geography of cases; where is it in my country, region, area? Is it in my town yet?	Number of cases: how many cases in country, town etc Detail is more important; where to avoid travelling to	Did I come into physical contact with [an infected person] while they might have been infectious?
Statistical information; numbers; facts	Developments in statistics; are numbers of cases increasing dramatically? Is it spreading more rapidly?	How did [the person I know/identify with] catch it? What mistake(s) did they make?
Scientific information; How does it spread? What can be done to contain it?	How fast is it spreading? Why are containment methods not working? Which are working better?	What should I do now? What can I do if professional healthcare cannot be accessed?
Passively receive information through news channels and official briefings	Actively seek information; share good information (Information push as well as pull)	Search for more granular information; exchange information with friends and colleagues.
Tendency to think the situation is being over-reported and the risk overemphasised	Situation is taken more seriously; more awareness of behaviour	How do I protect myself (including by social distancing and self-isolation); what to stockpile.

TABLE 20: *Difference in questions asked by far at-risk, near at-risk and actual at-risk groups, which differed as the outbreak came 'closer' to those being interviewed.*

### ***Technology Affordances of r/ebola***

One of the main advantages of an interactive discussion forum like r/ebola is that when many questions are highly specific and time-related (generally relating to cases that were in the news that day), which a pre-prepared Q&A would have difficulty anticipating, the forum users can quickly provide the answers needed. For example, the WHO FAQ on Ebola, at [www.who.int/csr/disease/ebola/faq-ebola/en/](http://www.who.int/csr/disease/ebola/faq-ebola/en/), approaches the question of how people become infected with Ebola as follows:

#### ***How do people become infected with the Ebola virus?***

*People become infected with Ebola either through contact with infected animals (usually following butchering, cooking or eating) or through direct contact with the bodily fluids of infected humans. Most cases are caused by human to human transmission which occurs when blood or other bodily fluids or secretions (stool, urine, saliva, semen) of infected people enters a healthy person's body through broken skin or mucous membranes.*

*Infection can also occur if the broken skin or the mucous membranes of a healthy person comes into contact with items or environments contaminated with bodily fluids from a contaminated person. These may include soiled clothing, bed linen, gloves, protective equipment and medical waste such as used hypodermic syringes.*

But is this sufficient information to answer the questions reported or observed during this study, which included (in no particular order): Can Ebola be caught from sitting next to someone on a bus; shaking hands with someone; an infected cleaner who has cleaned my house; sitting on a settee a person with

Ebola has sat on; being licked by a dog that has licked/eaten a dead body; dry vomit; a doorknob; money; a swimming pool; international mail; imported goods from infected regions; cell phones; and many others? It may be the case that to a scientist, the answer provided by the WHO FAQ does give sufficient information, for example the answer to the question, “can I catch Ebola from touching a doorknob?” may well be answered, “not unless someone who is infected with Ebola has very recently left bodily fluid on that doorknob, which has not yet dried, and the hand you touch it with has an open cut or abrasion and/or you immediately rub your eye with that hand”, but this may not be implicitly understood by someone without the scientific understanding to interpret the answer in that way.

On a discussion forum, someone may not only be able to answer whether one can catch Ebola from a doorknob but also elaborate on whether the type of doorknob might matter, compare the likelihood of catching it from a doorknob with catching it from other inanimate objects, the impact or not of wearing gloves, and how to open a door to minimize the risk of Ebola transmission.

In addition, the discussion forum can quickly answer how or why specific cases of infection – such as from Thomas Eric Duncan to two healthcare workers who treated him – have occurred, and why other speculative ones – such as from Kaci Hickox, a nurse who had treated Ebola patients but who showed no signs of disease and defied quarantine to go on a bike ride – were extremely unlikely to occur. Such forums represent a different solution to the information seeker than that most needed during Far at Risk, but it is no less valuable, particularly to those who are feeling Near at Risk and as the desire for personalised answers increases into the Real at Risk phase.

The technology affordances of reddit play well into the requirement to provide different types of platform. General information subreddits such as r/news,

r/science and r/worldnews, consisting of mainly link posts from a wide range of sources served the Far at Risk community well, but as Context Shift influenced a desire for more information, and for that information to be higher quality than speculative media reports, a new community formed. They migrated to a dedicated subreddit, r/ebola, where those who wanted to actively seek out information could do so more easily as it was collated into one place and not interspersed with news on other events. This new community then exhibited Community Shift. As the outbreak became more serious and cases were recorded in the U.S., the composition and needs of the community changed, but r/ebola adapted and responded through its ability to support questions and have them answered. New features did not need to be introduced, but the popularity and frequency of use of certain existing features changed, with the ability to ask questions in self-posts becoming more important. As the outbreak died down, a further Community Shift has been observed, as the behaviour has returned to Far at Risk, or perhaps evolved to Post-risk, depending on whether the Far should only denote geographic distance or can be expanded to include temporal.

How these differences in the communities using the forum at the different phases affects the characteristics and quality of the information the crowd can provide has not previously been studied and I will turn to this in the following chapter, Chapter 6. We also need to learn more about how the r/ebola community was formed, grew and was managed, and I will return to this in Chapter 8.

What does this tell us about the technology affordances that are most required during a PHEIC? Namely, this: if information sharing platforms that already exist when an outbreak emerges are likely to experience a Community Shift, in which the health-seeking and information sharing behaviour observed within their user community changes as the context changes, the platform

needs to be flexible enough to adapt to those changing needs, or else new platforms with different affordances more appropriate for the new contexts, and the different communities those contexts create, may have to be provided. During a public health emergency, neither the context nor the community may be able to be clearly delineated or easily bounded, however, particularly in a non-geographically bounded online space where users from Far at Risk, Near at Risk and Real at Risk (and eventually Post-at-Risk) communities may all be sharing information across the same platform simultaneously.

### ***Outbreak Subreddits: From Newsfeed to Community Space***

As sources of information, r/ebola (and r/zika) appear to be a significant departure from the characteristics of the type of health forums that have most frequently been studied in the academic literature. While chronic condition subreddits might be categorised as *Support* subreddits, PHEIC outbreak subreddits are, at different stages of the outbreak, *Information* subreddits and *Special Interest Community* subreddits; their moderators do not have direct experience of the health condition and the style of posting – more link posts than self-posts – has more in common with news subreddits than other health subreddits. For example, r/diabetes has very few link posts (none on its Front Page on 3<sup>rd</sup> March 2017) and approximately half its posters denote they have diabetes by using flair; the rare posts from users concerned about developing the condition tend to be from people in high risk groups, such as those with several family members who have it. Health forums appear to be predominantly for people who are living with a health condition, not those who are trying to avoid developing one. The NHS-supported HealthUnlocked, which has more than 500 active forums, has none discussing infectious diseases such as influenza, chickenpox, norovirus or measles and the forums dedicated to tuberculosis and HIV are aimed at patients and their families already living with the conditions. There is some discussion of how to avoid

catching sexually transmitted diseases in the Sexual Health Matters forum, but this is generally in response to people who are asking about the likely outcome of risky behaviour and are being warned to be more careful in future.

Seeking information from a condition-specific forum in a 'Far at Risk' and 'Near at Risk' context does not seem to be common to other forums dedicated to health conditions, and there is therefore very little with which to compare the data from these subreddits. The small amount of data available from the handful of WAG interviewees who were briefly Real at Risk does, however, suggest that a Context Shift to Real at Risk, resulting from people having first-hand experience of Ebola and having to live with it rather than only attempting to avoid it, would also result in a Community Shift in which the community became more inward-looking, shrinking the size of the crowd its members want to consult, and becoming more supportive of one another, sharing more personal experiences and advice. This suggests that such a Context Shift would see any such subreddits dedicated to the outbreak aligning more with the chronic disease subreddits than the newsfeed subreddit of r/worldnews, with which r/ebola currently has more in common.

## **5.8 Conclusions**

There are clearly important lessons to be learned from the questions the WAG group and the r/ebola users had during the Ebola crisis. Despite the limitations of a small sample size sourced from the socioeconomic elite of the affected countries, the study sheds light on what questions and concerns those at risk may have at different stages of an outbreak, how they ask them and the type of platforms through which those might best be answered. As some questions that arose during the Ebola outbreak are likely to be generic to other outbreaks, as are the events that triggered them, this may assist public health information provision during future health emergencies. If questions can be anticipated, appropriate answers can be formulated more quickly and

information can be presented in such a way that such questions are less likely to arise. Symptoms can be contextualised to whether ‘gravely ill’ means the sufferer is likely to be walking around or bedridden, and the Basic Reproductive Rate ( $R_0$ ) explained whenever it is used. Provision needs to be made for people who do not understand the scientific and medical terms being discussed to be able to ask for clarification where needed.

This chapter has helped to provide a clearer understanding of what questions people asked during the Ebola outbreak, and how r/ebola was able to provide a forum in which they could be answered. The next consideration, however, is how good were the answers received from r/ebola and how did users judge the credibility of those answers? This will be explored in depth in the following chapter.

## 6: INFORMATION RETURN: ANSWERS FROM REDDIT

---

### 6.1 Introduction

In the previous chapter, I investigated the types of information people may want during a PHEIC and showed how reddit's characteristics make it well-placed to provide that information. In this chapter, I turn to how reddit's characteristics influence the credibility and accuracy of information. As not all information on reddit may be equally accurate or equally trustworthy, can signposts help users to find the items that are more likely to be high quality and steer them away from those of lower quality?

To do this, I will examine the surface credibility of r/ebola; the message credibility of its posts and comments; and the quality of the information found in the messages it signposts as credible.

### 6.2 The relationship between trust and quality

The definition of quality I use in this chapter was developed for use in the surveys given to the doctors who took part in study Phase II-B, to help them assess the information found on r/ebola. The survey document used in study Phase II-B can be found at Appendix IV. This definition is:

*“Information which is accurate, and is in-line with what would be expected from a qualified medical or public health practitioner; which a qualified medical practitioner would consider to be sensible and appropriate in light of the question asked; and gives information which is unbiased and does not present one view as the only available option if more than one is available.”*

The definition of ‘trust’ is taken from the final stage of content evaluation in Wathen & Burkell’s model:

*“Am I ready to believe this information? Am I ready to act on it?”*

The distinction between information users *do* trust and information they *should* trust in online health information is important. Medical information on the Internet is mostly free, easy to access, available from all over the world and able to be extremely up-to-date but it is also unregulated (Impicciatore et al 1997) and may not have been subject to the same rigorous editorial or quality checks as professional medical publications and journals that ensure errors are weeded out (Akerkar and Bichile, 2004).

People who are looking for information on a subject they know little about may find it difficult to tell good information from bad. This may be especially true during a public health emergency involving a new disease of which no-one has much experience. Discussion forum users may be legitimately concerned about whether it is possible for a virus to become airborne, or whether it is possible to catch the disease from passengers on a crowded bus or from handling money – even if experts know such eventualities are unlikely – and will need help in assessing the quality of information they receive. Each comment posted in reply will need to be assessed for credibility and accuracy, as will each forum which a question might be asked.

Trusting online health information carries a high risk, as the consequences of following erroneous advice may, in extreme circumstances, be life-threatening (Luo and Najdawi, 2004), making signposts such as established branding and third party accreditation seals more important on health sites than on general consumer ones. While many studies have looked at why people use the Internet to find health information and what makes them trust the information

they find (Bernhardt and Felter, 2004; Sillence et al, 2013; Kim, 2016), few then determine if trusted information is in fact accurate.

Health information found on the Internet is known to vary widely in quality. An early study by Impicciatore (1997), on web-based advice for the home management of feverish children, found that information was often incomplete or biased, though very little (only around 5%) was measurably inaccurate. Studies by Suarez-Almazor et al (2001) on rheumatoid arthritis, Air et al (2007) on thyroid cancers and Whitelaw et al (2014) on birth choices following Caesarian section also found little inaccurate information but a broad range of completeness and some inherent biases in how information was presented. The quality of information can also differ depending on the health condition, particularly where appropriate treatment may be more subjective and there is less medical consensus – a study of online slimming information (Miles et al, 2000), considered nearly 90% of information to be inaccurate to some degree.

These studies cited above, and dozens of others like them, tend to focus on information relating to either a specific medical condition or disease (e.g. rheumatoid arthritis, thyroid cancer) rather than the characteristics of the website(s) on which it is found, however. I have not been able to find studies that focus on how the characteristics of a specific website or type of website might help the user to trust the information they find, other than those that point to sites owned by professional health brands. Instead, ‘the Internet’ is most often treated as an amorphous whole. The existing literature makes little distinction between the official websites of professional health bodies and personal blogs when identifying that quality varies widely, and there has been little examination of how discussion forums compare with other types of web platform format.

Mentions of online discussion forums in the existing literature are limited to observations that they appear to be underused compared to other types of website (Dahlen et al, 2013; Whitehal et al, 2014), even though a study by Marco et al (2006, p1041) into an ‘ask the expert’ site for people living with HIV concluded that there is “a great demand for this type of ‘ask the expert’ Internet service, at least for AIDS and hepatitis”. Anonymity, free access and immediate answers were considered key factors in the platform’s success but in general, most studies on health information focus only on the content – the message characteristics – of the website. This misses a valuable opportunity to determine if some channel characteristics, such as the voting system on reddit that indicates how other reddit users have rated the information, help to highlight better quality comments and posts.

Eysenbach and Köhler (2002) actively excluded ‘newsgroups and email groups’ from their systematic evaluation of online health information, without giving a reason why. Lawrentschuk et al (2012) included video sites in their study of oncology information online, but only considered how their characteristics affected the information they presented, not how sites allowing user-generated material to be posted can stimulate interaction or manage the quality of the material they present. San Norberto et al (2011) focused on how easily websites relating to aortic aneurysm (*aneurisma aorta*) and aortic endoprosthesis (*endoprotesis aorta*) could be understood, but did not explore how users might discuss the information in a two-way exchange. Groselj (2014) considered the linkages to and from social media and blogging sites to other types of website, but did not determine the quality of any of the websites included in the study. There is a lack of evidence on how websites are able to present high quality information, rather than simply whether or not they do.

The few studies to acknowledge that discussion sites may have specific characteristics have done little more than note this and state that it has been

outside their scope to explore further (e.g. Whitelaw et al, 2014). None examine what might help to influence or signpost quality then go on to check if it indeed does. Health information found on the Internet is not necessarily less accurate or complete than in other media such as broadcast television or magazines, nor does the existence of poor information automatically translate into poor health outcomes (Eysenbach and Köhler, 2002), but how users navigate the information provided is poorly understood. There has been very little investigation of the consequences – if any – of poor information being posted, nor of how it is received by the reader. This is a research gap this chapter seeks to address.

Accreditation of some kind, such as the World Health Organization supported HONcode (Health On the Net Foundation, [www.hon.ch/HONcode/](http://www.hon.ch/HONcode/)), accreditation by the American Medical Association, or URAC (originally, Utilization Review Accreditation Commission, [www.urac.org](http://www.urac.org)) is often used as an indication of likely quality (Akerkar and Bichile 2004; Luo and Najdawi, 2004; Lawentschuk, 2012) but the studies do not investigate how the quality of sites that do and do not display such accreditation compare.

In the next sections, as I assess the credibility and the accuracy of information on r/ebola, I also consider the importance of assessing these qualities together. A platform with affordances that improve the credibility of information also needs to be able to apply those credibility signposts to the most deserving information.

### **6.3 Credibility on r/ebola**

Phase I of my study suggested that during the early phases of an outbreak, the most important component of a message is the source. In the WAG interviews, this was much more influential than the characteristics of the channel through

which the information was received. In most cases, WAG interviewees could not recall how they had first received information, just what category of sender it had come from (informal, official information or professional media, etc). As information seeking became more active, those seeking information turned to trusted channels such as WHO and BBC as their desire for more information grew, and demonstrated scepticism of information that came from social media or sensationalist newspapers.

Users of reddit put significant store in the originator of material posted on reddit, whether in link posts or self-posts, as well as valuing the brand of reddit itself. This provides signposts to surface credibility that help to build confidence in the site itself and help users to evaluate message credibility. Users are also shown, by the voting activity, how other members of the community evaluate the message.

Wathen and Burkell's model was designed to assess a single website, but reddit has a series of layers, each of which creates a space where knowledge may be found and constitutes a collection of information and of people who may hold information they are yet to contribute. Each layer needs to achieve its own surface credibility for the user to progress through the layers to the actual message, which will then be assessed for message credibility. These layers are:

- **reddit** itself
- the **subreddit(s)** the knowledge seeker chooses to consult or post on in their attempt to answer their question
- the **posts** within a subreddit which may or may not answer their question and which, if they don't, may require the knowledge seeker to make a new post, and
- the **comments** made on the chosen post(s).

If surface credibility is achieved at each of these stages, the user will arrive at comments containing the answer they were seeking. They can then assess the message credibility of these comments, helped by additional signposts that show them which information has been considered credible by the wider reddit community, whose Collective Intelligence may be able to guide and improve their own health literacy.

### ***Layer 1: Surface credibility of reddit***

Surface credibility depends on the look and feel of the website, how easy it is to navigate and how quickly and easily the user feels they can get what they want. Here, reddit has many advantages. It is a professionally run website on which the basic design, architecture and processes are set, into which the individual subreddits can 'plug and go'. It is a trusted brand, with a high proportion of users within the UK population (Morse, 2015). In the pre- and early days of the Far-at-Risk stage, when reddit users first encounter information on Ebola (or a future disease outbreak) through general subreddits such as r/news, r/worldnews and the Front Page, they are likely to consider this information credible.

### ***Layer 2: Surface credibility of subreddits***

A user who decides to actively seek out information after encountering news passively on the Front Page, or on r/news, may do this by determining if there is a subreddit dedicated to the subject: a search on 'Ebola' within reddit will bring up a list of any subreddits that exist. Signposts attached to the information returned by this search provide (or challenge) surface credibility by giving indications as to the likely positioning and popularity of each of the subreddits returned by the search.

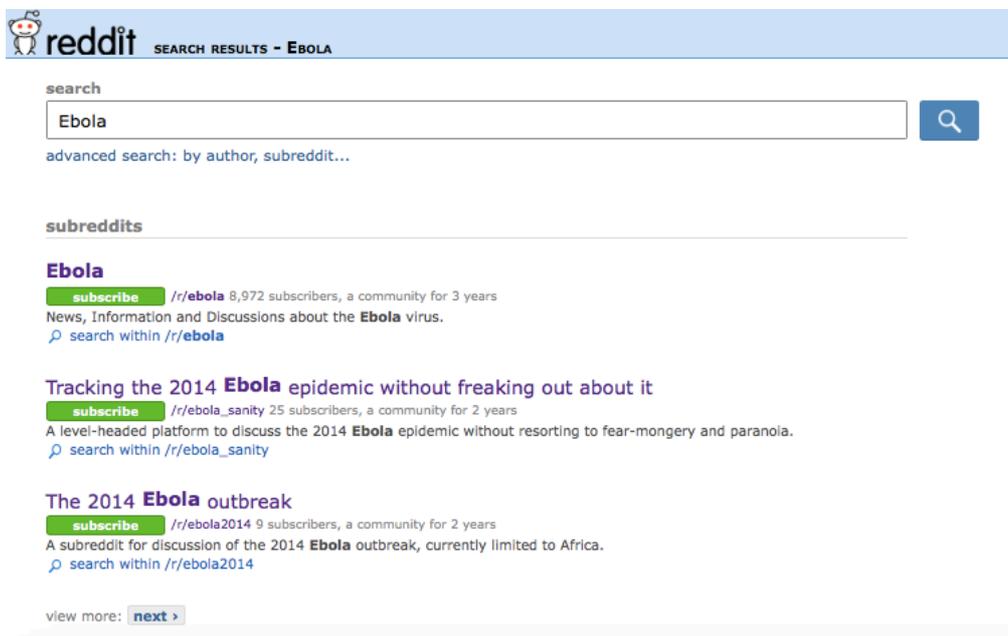


FIG 29: A search for Ebola-related subreddits gives indications of the subreddit's positioning, the number of subscribers and the length of time it has been established.

A search undertaken on 30<sup>th</sup> November 2016 returned the results shown in FIG 29, above. This shows that [r/ebola](#) had more than 9,000 subscribers, whereas [r/ebola\\_sanity](#) had 25, and [r/ebola2014](#) just nine. While this is an indication of popularity rather than quality *per se*, it is not unreasonable to assume that more popular subreddits are also providing more relevant information to their users. The strapline beneath the subreddit title gives a brief indication of the forum's positioning and shows that [r/ebola\\_sanity](#), while it might be small, aims to provide level-headed information on the Ebola outbreak; a user might choose to click through to this instead of the more popular forum. If they do so, however, they will find that it is in fact a redundant subreddit, which contains just 11 posts, all but two of which were made by the single moderator, on 8<sup>th</sup> and 9<sup>th</sup> October 2014. Further clicking through to the moderator's homepage shows that they have not made a single post to reddit

in more than a year. r/ebola2014 is an empty subreddit containing no posts at all. One might therefore assume that the typical information seeker would choose to proceed to r/ebola, the most popular (by far) Ebola subreddit, which positions itself as ‘a subreddit for news, information and discussions about the 2014 Ebola outbreak which originated in West Africa’. The subscriber numbers indicate that this subreddit provides the largest crowd from which an answer might be sought.

The search results also return posts, as well as subreddits, and this shows that significant discussion on Ebola has also taken place in r/science, r/IAmA and other subreddits covering broader topics.



The screenshot shows a list of four search results for the term 'Ebola' on Reddit. At the top, it says 'posts' and 'sorted by: relevance' with a dropdown arrow, and 'links from: all time' with a dropdown arrow. The first result is a 'Discussion: First Case of Ebola in the United States is diagnosed in the Dallas (Mega Thread)' with 411 points and 1,513 comments, submitted 2 years ago by user 'lifeless2011' to r/ebola. The second result is a 'Science AMA Series: Ask Your Questions About Ebola' with 3,627 points and 2,649 comments, submitted 2 years ago by user 'nallen' to r/science. The third result is a 'PSA - Want to help fight Ebola? Get a flu shot.' with 285 points and 294 comments, submitted 2 years ago by user 'midnitewarrior' to r/ebola. The fourth result is 'IamA 14yr old Ebola survivor in remote Liberia' with 3,578 points and 2,547 comments, submitted 1 year ago by user 'reese015' to r/IAmA. Each result includes a 'more' link at the bottom.

FIG 30: Significant discussion on Ebola has also taken place outside Ebola specific subreddits, such as r/science and r/IAmA, which may direct a user to seek information there instead.

The user could decide to look for answers in any of these, instead: the crowd will be larger but a lower percentage of it may be experts in the topic of interest. This could require the information seeker to make a value judgement on whether it is better to post their question to a large crowd of whom only a small proportion may be able to answer, or a small crowd of which a higher proportion may be likely to answer but a valuable technology affordance of reddit is cross-posting – being able to post the same question in more than one subreddit simultaneously. Examination of this was outside the scope of this study, but a useful future research project would be to examine the answers received by cross-posted questions and if these differ depending on the subreddit from which the answers are received. As previous studies (Bernhardt and Felter, 2004; Sillence et al, 2013) and Chapter 5 of this thesis have shown, health information seekers often prefer to compare information from different sources and find comfort in consensus, cross-posting may be an important affordance of the reddit platform. Users can, of course, consult all existing subreddits: using one does not exclude them from also using the others.

### ***Layer 3: Surface credibility of posts***

Assuming the information seeker decides to investigate r/ebola further, once inside the subreddit, there are additional markers that can guide them towards the posts that are more likely to contain high quality information than others.

The reddit ordering system is itself an indicator of credibility, indicating which posts have been more favourably received by the subreddit community. Beyond this, each post has five separate, but complementary signposts: indication of link- or self-post [1], flair [2], two source indicators – of the original material [3] and the user who posted it [4] – and voting activity [5]. The first four relate to the source characteristics of the post, providing credibility through the reputation or expertise of the information's originator,

and the fifth relates to receiver characteristics, indicating how well or not the information has been received by other reddit users.

Wathen & Burkell's model considered only the receiver characteristics of an individual platform user, however, not how a user *community* receives information and how this might influence credibility within that community. I believe this is a valid concept to consider. The receiver characteristic of *being a redditor* not only adds to the credibility of information found on reddit for that user, but is also likely to make them value the judgements of other reddit users, so that content evaluation by the user community may guide an individual user's own evaluation.



FIG 31: Posts on r/ebola display a number of signposts to potential quality, including whether or not the information is contained in a link-post or a self-post; flair, which labels the post against certain criteria such as science/medicine, EduSig (education special interest groups), media etc; the original source of the information in the case of link posts, such as the journal in which it appeared or the media source reporting on it; which reddit user has posted it; and the score it has received from the reddit community.

The first signpost I will consider is whether the post is a link-post, which leads to content hosted on an external website, or a self-post, which is usually a question asked or opinion stated by a reddit user with no link to external material. In general on reddit, link posts are considered to contain more credible information than self-posts, which can contain personal opinions and/or soapboxing. They are strongly discouraged on some subreddits (including r/ebola and r/science); others limit them to certain days of the week.

The second indicator is flair, shown in the tab to the left of the link- or self-post indicator. On r/ebola, 27 different flairs are available, including 'science/medical' for posts from official scientific and medical sources, flairs for international and governmental health organisations such as CDC and WHO, and 'media' or 'speculative', for sources that may be likely to put sensationalism ahead of scientific accuracy. A flair can be assigned by the poster at the time of submission, or afterwards by a forum moderator. Users can search for material by filtering in only posts from 'science/medical' sources, for instance, while filtering out any media or speculative posts.

The third signpost is the indicator of who made the post. This appears just under the post's title and clicks through to the user's home page. This provides a sense of their views, attitudes, and karma scores (how well their previous posts have been received by the reddit community). Some subreddits assign flair to reddit users as well as to link sources, indicating whether they are, for example, medically qualified (for r/docs), have a degree-level education or above in a certain field (r/science), or give details about their health condition (r/diabetes, r/cancer), which indicate their likely level of experience. r/ebola considered implementing flair for medics and others working on the 'frontline' of the outbreak but had insufficient time to determine a reliable method of checking credentials, though this could be considered in future disease outbreak subreddits.

The fourth indicator on the post is the url of the website where the information was originally published, such as a scientific journal, the press release pages of professional health organisations, or external media organisations. This link takes the user to a history of all reddit posts linking to that organisation/source and provides an overview of the originator's position – for instance, is all the information pushing one political view or conspiracy theory, or is it all evidence-based science? Self-posts list 'self-ebola' as the source, and link back to the subreddit homepage. There can be many layers to the source of a link post – for instance, the post may link to a news item from a media source, which in turn quotes a scientific study published in a peer-reviewed journal written by academic researchers, who have their own website at their academic institution or organisation – but the indicator on the post at least gives a starting point from which the reader can begin to evaluate its likely quality.

Finally, the last indicator of quality is the voting score the post has received, and what percentage of votes were upvotes. This indicates how well it was received by the reddit community and, if the voting system does indeed indicate quality rather than just popularity – as I will test later in this chapter – this offers a particularly valuable technology affordance that is not present on Google or other search engines.

San Norberto et al (2011) argued that a position among the first web pages appearing on a search engine does not guarantee that the information provided is relevant or accurate, nor that it will be easily comprehensible to a non-medical expert – but reddit is different. While the default setting of internal searches works on algorithms set to detect relevance, the results returned can be ordered by 'Top', displaying the posts according to how they were voted on. If a post or comment has been rated highly by the collective votes of the r/ebola crowd, this should be a strong indication to members of that community that they, too, might do well to consider it credible.

#### ***Layer 4: Surface credibility of comments***

Whether the information seeker decides to consult existing posts and their comments, or generate a new self-post and wait for comments to be returned, the surface credibility indicators of any one comment are essentially a subset of the markers for posts (the username of the redditor who made it; how many upvotes-downvotes it has received) plus how well-written and clearly expressed it is (Tan et al, 2016).

By the time a user has navigated all four layers, and assessed the surface credibility of each, it is fair to assume that the process confers at least some message credibility on the information found at the end of it. The signposts provided by reddit help to guide information seekers not only to information they want, but also that they can be confident of trusting.

It needs to be acknowledged that for most reddit users, the processes described above are almost certainly unconscious. They also require a knowledge of, and familiarity with, reddit that new users may not have. If it was considered advantageous to direct new and inexperienced users towards such a forum in future, ways to explain the intricacies of navigating surface credibility may need to be considered. On the other hand, as these markers are consistent across all subreddits, a user familiar with one reddit forum – such as r/news – would be able to easily and seamlessly begin to use another, whether they were aware of those processes or not.

#### **6.4 Searching for answers – the typical information-seeking journey**

Imagine then, the journey of a reddit user who has seen news stories on r/worldnews that Ebola may become airborne, and is looking to answer the question, “How likely is this to happen?”. One way to find an answer would be to search, first for the term ‘Ebola’, which will return a list of all the

subreddits relevant to the term. On the basis of the surface credibility indicators discussed above, I will assume they choose to visit r/ebola.

Once 'in' r/ebola, the information seeker would either find posts on its home page discussing the news story, and could assess them on the post credibility markers discussed above, or, if there was no discussion, might then decide to search on the term 'airborne' within the subreddit (perhaps noting that if the story was not reaching the front page, this might indicate a low hotscore and therefore low credibility). On 10<sup>th</sup> August 2016, such a search returned the results shown in FIG 32, on the following page.

How this information might be interpreted, however, may differ depending on how experienced the information seeker is with reddit and if they are aware of how the surface credibility indicators work. An experienced reddit user may take the results cautiously: all but two of the posts are self-posts, which r/ebola's guidelines discourage. Indeed, at the point this search was made, the r/ebola front page displayed only link posts, suggesting that text posts are unusual. The user may choose to read the posts anyway – and if they did, they should come away with a sense that the likelihood of Ebola virus becoming airborne is highly unlikely, reassured that it is certainly not capable of spreading through airborne transmission at present – or they may instead choose to narrow their search to posts which have been 'flaired' as being from reputable scientific and medical sources (the science/medicine flair).

In r/ebola, this does require knowledge of how the subreddit's filtering system works, as it was only partly set up at the height of the outbreak by a moderator who then left the forum, and is not as user-friendly as those found on other forums such as r/technology. The user has to search on flair:science/medicine (or whichever flair is required) using the internal r/ebola search box.

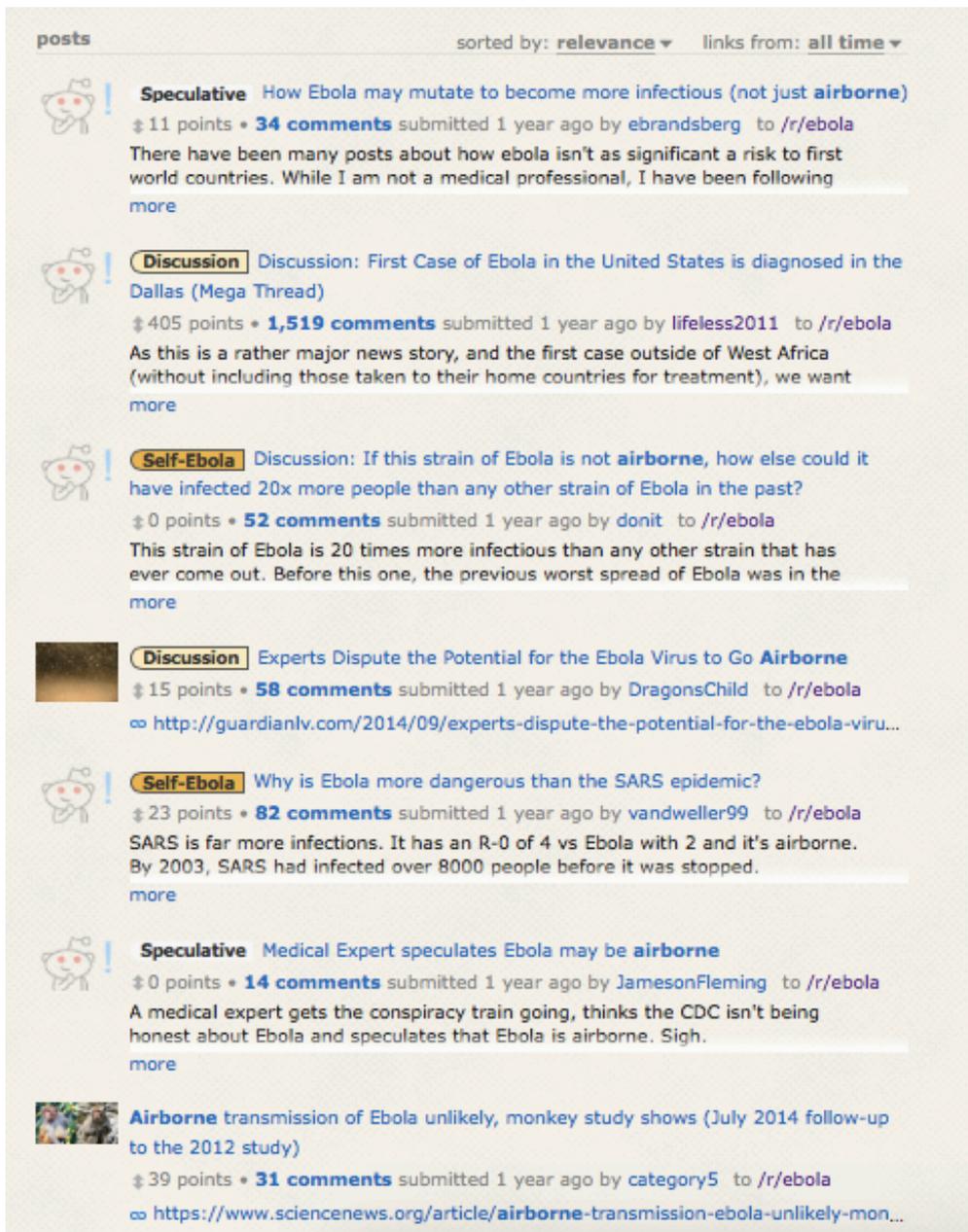


FIG 32: Results returned for a search inside r/ebola on 'airborne'.

The example below shows a search carried out within r/ebola to filter only stories on airborne ebola from science/medicine sources. The search returns a single result, which takes a clear position that the Ebola virus will not go airborne. Searches on flair:WHO and flair:CDC + airborne return no results.

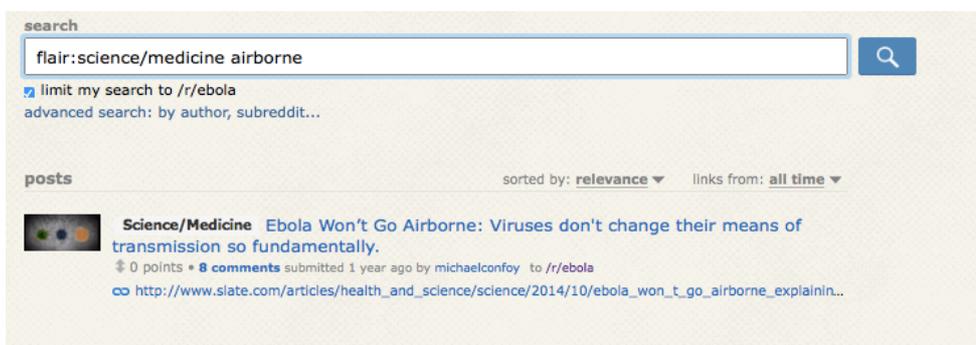


FIG 33: *The only story about airborne transmission of Ebola from a source flaired as science/medical takes a clear position this is not likely to happen.*

Following this path should, therefore, provide the reddit user with clear information that the Ebola virus is not, and is not likely to become, airborne.

Should the user want to check further, r/science, which only allows link posts to peer-reviewed scientific journals, contains only three posts about airborne Ebola, all of which would also help convince an information seeker that such a mutation is unlikely. The fact that the same conclusion can be drawn from two different subreddits may itself act as a quality assurance mechanism.

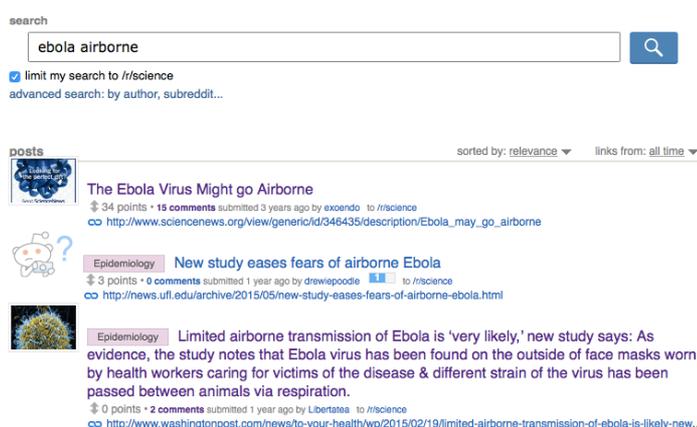


FIG 34: *Only three posts about Ebola becoming airborne are found on r/science, one from Science News, one from a university and one from the Washington Post.*

search

ebola airborne

limit my search to /r/worldnews  
advanced search: by author, subreddit...

posts sorted by: **relevance** links from: **all time**

**Misleading Edit In Title** Serious Ebola Outbreak in West Africa, Rumored to be Airborne.  
 ‡ 1,447 points • **584 comments** submitted 2 years ago by Alstroph to /r/worldnews  
 ∞ <http://www.independent.co.uk/news/world/africa/panic-as-deadly-ebola-virus-spreads-acr...>

Limited airborne transmission of Ebola is 'very likely,' new analysis says  
 ‡ 36 points • **5 comments** submitted 1 year ago by Sybles to /r/worldnews  
 ∞ <http://www.washingtonpost.com/news/to-your-health/wp/2015/02/19/limited-airborne-tra...>

**Ebola** (In response to Michael T. Osterholm), Airborne Ebola Is Extremely Unlikely -- Anthony Fauci (NIAID)  
 ‡ 19 points • **2 comments** submitted 1 year ago by princetonwu to /r/worldnews  
 ∞ <http://time.com/3342305/airbone-ebola-not-happening/>

**Ebola** UN Ebola chief raises 'nightmare' that virus could become airborne.  
 ‡ 12 points • **13 comments** submitted 1 year ago by Arbitelle to /r/worldnews  
 ∞ <http://www.dailymail.co.uk/news/article-2778022/UN-Ebola-chief-raises-nightmare-prospe...>

**Ebola** CDC concerned over airborne Ebola Zaire - news from 2012  
 ‡ 2 points • **3 comments** submitted 1 year ago by Kancho\_Ninja to /r/worldnews  
 ∞ <http://www.bbc.com/news/science-environment-20341423>

**Ebola** US scientist: Ebola unlikely to become airborne  
 ‡ 0 points • **3 comments** submitted 1 year ago by Topsy\_Fox to /r/worldnews  
 ∞ [http://hosted.ap.org/dynamic/stories/U/US\\_MED\\_EBOLA\\_MUTATION?SITE=AP&SECTION=...](http://hosted.ap.org/dynamic/stories/U/US_MED_EBOLA_MUTATION?SITE=AP&SECTION=...)

**Ebola** Scientists warn Ebola transmission may be airborne; urge full respirators for frontline health workers  
 ‡ 0 points • **9 comments** submitted 1 year ago by gari-soflo to /r/worldnews  
 ∞ [http://www.naturalnews.com/046986\\_Ebola\\_airborne\\_transmission\\_respirators.html#ixz...](http://www.naturalnews.com/046986_Ebola_airborne_transmission_respirators.html#ixz...)

FIG 35: The same search, on 'ebola airborne' on r/worldnews flairs one post as 'Misleading Edit in Title'. One of the most speculative posts – 'UN Ebola chief raises 'nightmare' that Ebola could become airborne', is from the Daily Mail, a media source blacklisted by r/ebola due to its misleading and oversensational coverage. Note that r/worldnews has an 'Ebola' flair, a system introduced to enable all coverage of highly topical titles to be filtered for, or filtered out of, its front page.

The above process explains one way in which a reddit user can attempt to answer their question: by searching for pieces of information already lodged within the reddit collection, and absorbing them to help construct an answer. The other way is to find the most appropriate subreddit where others are discussing the topic, and to ask the question hoping that people will answer and that the community will help to sort the better answers from the poorer ones through the voting system.

The voting system is clearly an important signpost to message credibility. The next question to ask is does it also denote message quality? To what extent do highly upvoted posts and comments represent better quality information?

## **6.5 Quality of information on reddit health forums**

While reddit is undoubtedly capable of displaying a wide spectrum of information quality, I would challenge that it is unregulated. The credibility signposts are an informal, but very effective indication that regulation is taking place, and understanding how this process works may help doctors assess whether certain types of website (and certain specific websites) are more likely to contain accurate information than others. As the existing literature suggests that inaccuracy in online health information is rare (Esquivel et al, 2006; Corocco et al, 2002; Nölke et al, 2015) but that bias or incompleteness is more likely (Impicciatore et al, 1997; Whitelaw et al, 2014), it is useful to know the extent to which reddit's structure might mitigate this.

Quality in information systems in general has been defined as having many interrelated factors, including that it should be personalized, complete, relevant, easy to understand and secure (DeLone and McLean, 1992 and 2003). The discussion forum nature of reddit, that enables communities of subreddits to be formed through which users can ask questions and receive

replies, fulfils the requirement for information to be personalized; the fact that so many people already use it suggest that it is easy to understand and navigate. This still raises the question, however, of whether it is complete and relevant, and by extension accurate, and so whether the system does enable Collective (health) Intelligence to emerge from wise crowds of users.

### **6.5 Phase II-A: Investigation of information quality in online health forums (pilot study)**

Early in the study design process, I designed a pilot study to investigate how I might test likely accuracy of information in online discussion forums by asking doctors to rate it. As the pilot was devised before the severity of the West Africa Ebola outbreak became apparent, or that it would offer such a suitable case study, the intended focus at the time was on health information in general. The study sample was small, as it is hard to find doctors with enough time to complete a detailed survey, and not enough data was returned to make statistically significant comparisons, but it nonetheless provided valuable insights into how the main study should be conducted.

#### ***Methodology***

I aimed to determine the general quality of health information across three online forums (reddit [[www.reddit.com](http://www.reddit.com)], Mumsnet [[www.mumsnet.com](http://www.mumsnet.com)], and Patient [[www.patient.co.uk](http://www.patient.co.uk)]) by asking doctors and non-medically qualified members of the public to rate information found there. The intention of this was to determine an absolute score from the doctors (i.e. was the information ‘good’ or ‘bad’) and to see if the public could identify information considered ‘bad’ by doctors by rating it accordingly. Mumsnet and Patient have no community voting structure nor indication of whether posts and answers have been favourably received, and so it is not possible to tell how information

posted on those forums has been assessed by its users; on reddit, where the community votes, these votes were also compared with the scores given by the doctors and the public (though the numbers are too low to provide statistically significant confidence).

Assessors were asked to rate information found in the discussion threads according to five criteria: accuracy, completeness, how sensible they considered the replies to be, how they thought the questioner would act, and how useful they thought the questioner would find the replies (links to the survey questionnaires can be found in Appendix II), giving a highest score of 1 and a lowest of 5. The intention of this was to determine if respondents who considered information to be inaccurate also thought that it could cause harm, as the two are not necessarily the same. The full methodology for selection of the health information to be assessed and the study participants is described in Chapter 3. Full results have been published in the *Journal of Medical Internet Research* (Cole et al, 2016).

### **Results**

In all, 79 completed assessments were returned by 17 individuals (eight medical doctors, coded M001-M008, and nine members of the public, coded P001-P009). When the ratings awarded were analysed, and arranged into five overall rating bands, (highest 5-8 assessed across five criteria; good 9-12; middling 13-17; poor 18-21; and lowest 22-25), the lowest available rating band was awarded only twice: assessor M005 scored Q21 at 23 out of a possible worst score of 25 (5 for all 5 criteria) and assessor M001 scored Q18 at 24. On assessments of individual criteria for each question, the lowest possible score of 5 was awarded only 11 times out of a possible 390, whereas the highest was awarded 53 times. Across all five criteria for all 79 assessments, 13 of the assessments placed discussion threads, over all five criteria, in the highest possible score band, 27 in the second highest, 31 in the

middle band, 5 in the second lowest, and two placed a discussion thread in the lowest rating band. This suggests that health threads on Internet discussion forum websites are more likely than not, by a factor of approximately 5:1, to contain information of high or reasonably high quality when considered across all criteria (40 assessments placed the discussion thread in the highest two bands for overall thread assessment: just 7 in the lowest two) and by a factor of 4:1 when assessed on individual criteria (203 scores of 1 or 2 were awarded out of a possible 390, 52 scores of 4 or 5). Extremely poor information is rare (11 examples out of 390 = 2.8%), less than the percentage of around 5-7% found in previous studies (Impicciatore et al, 1997; Biermann et al, 1999; Eysenbach and Köhler, 2002; Eysenbach et al, 2002). This suggests that discussion forum websites may be a useful platform through which people can ask health-related questions and receive answers of acceptable quality. The assessment scores given are shown in FIGS 36-38 and Appendix III.

Assessors tended to agree on which discussion threads contained good quality information but what constituted poor quality information appears to be more subjective. Not all respondents who rated information as poor agreed that the original questioner would have been led to act inappropriately based on the information presented.

It is also interesting to note that there are some differences between the assessments of doctors and non-doctors. In general, the doctors tended to give higher scores for accuracy (in FIG 37, 54% rated the discussion threads as entirely or mostly accurate) than the non-doctors (who rated only 25% of the threads as entirely or mostly accurate, see FIG 38), and doctors were also more likely to give threads the lowest rating – all 11 instances of the lowest score were awarded by doctors. The non-doctors were more likely to give mid-range scores, suggesting that they may be less confident in their ability to assess information and more cautious in making strong assessments.

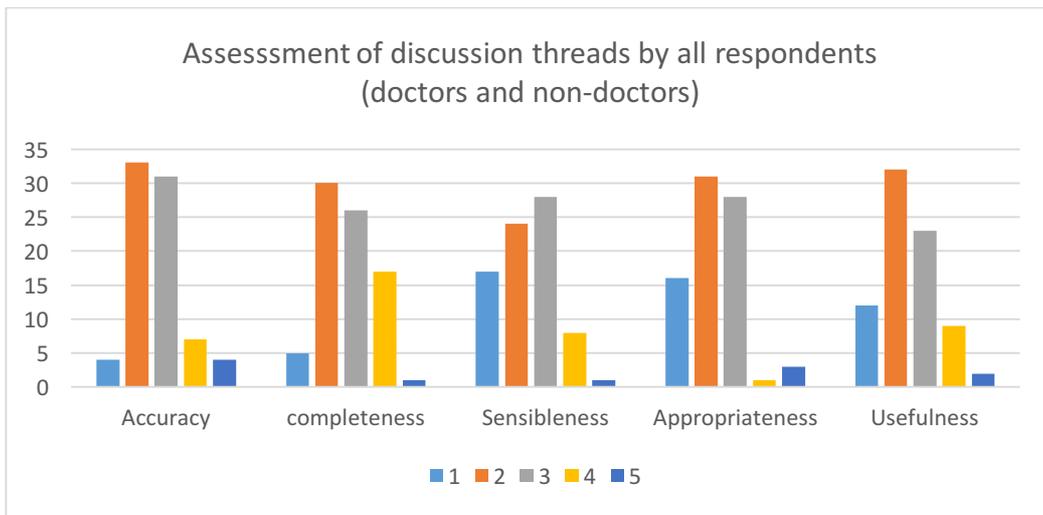


FIG 36: A higher percentage of respondents (indicated by the light blue column, far left) rated information as highly sensible, appropriate and useful than rated it to be highly complete or accurate. Few respondents (indicated by yellow and darker blue columns) thought information would lead to inappropriate action. Low quality information may not lead to actual harm.

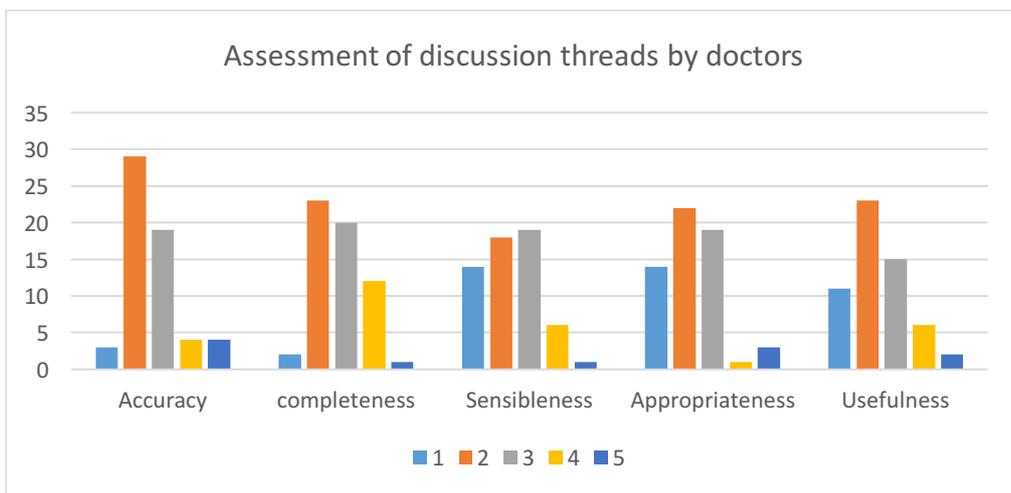


FIG 37: Doctors considered a significant proportion of the information to be mostly accurate and mostly complete (the dark orange column), and a good proportion to be very sensible and very appropriate (the dark blue column).

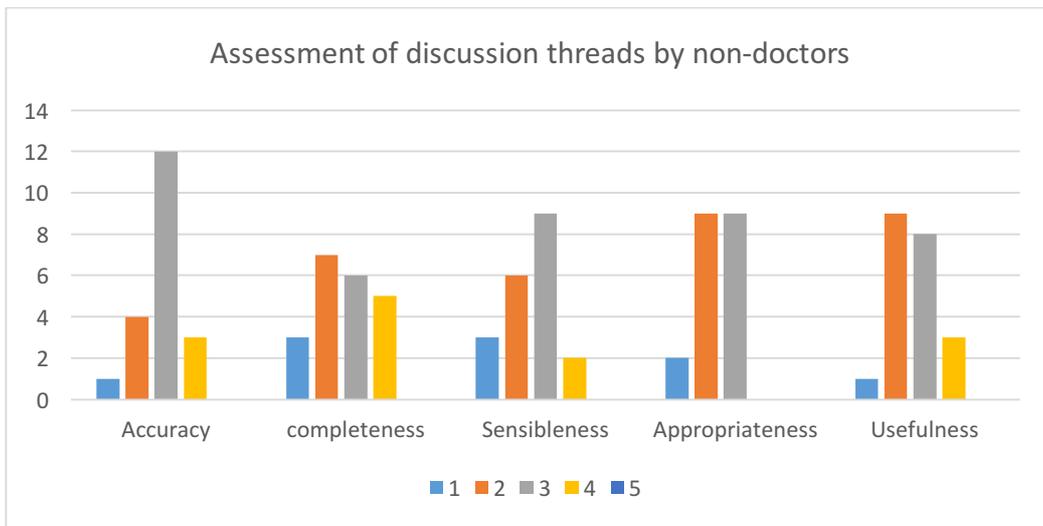


FIG 38: *The non-doctor respondents were more likely to award mid-range scores (indicated by the grey columns), suggesting they were less confident in their ability to identify very good or very poor information.*

More important than the results themselves, however, were the insights the pilot study provided into the nature of information on the Internet and how it is received, which informed the rest of the study.

Firstly, there was not much evidence of poor information. Threads contained very little information which I considered to be bad (or which, on reddit, had been heavily downvoted). I found this curious, and made exploratory posts on five reddit health forums (for diabetes, HIV/AIDS, asthma, multiple sclerosis and cancer) plus forums for healthcare professionals (r/askdocs, r/health2, r/publichealth, r/doctors and r/nursing) to ask about users' experiences in a forerunner of what became Phase III of the study. I received 22 comments on the posts I made, with replies reporting that poor information is rarely posted, and anything that is posted is often swiftly removed by forum moderators.

Offered examples of 'bad' information were mostly alternate remedies or dietary advice aimed at slimmers and bodybuilders. I was usually directed to an entire subreddit (e.g. r/homeopathy/), rather than a single discussion or

comment. This highlighted that the characteristics of a forum, such as its positioning and declared biases can have an influence on the information posted there, and informed my decision to focus within selected subreddits (such as r/ebola) in later phases, rather than across reddit entirely. I also decided to focus specifically on subreddits which signposted themselves as being aligned with mainstream medical and scientific approaches, as one might expect information on natural remedies to be received differently in r/homeopathy than in r/cancer, and as this study intends to determine the degree to which discussion forums might be helpful to professional health organisations, it was important to focus on forums likely to be aligned with their position (n.b. a search on 3<sup>rd</sup> March 2017 returned no posts containing the word ‘cancer’ on r/homeopathy and only one within r/aromatherapy, which pointed out that aromatherapy cannot cure it). It was also noticeable that the same discussion thread could be rated very differently by different assessors, even where those assessors were medically qualified. The discussion thread for Q21 (about whether it was safe to expose a baby to a child who had chickenpox) received a ‘very poor’ rating overall from assessor M005, but a ‘good’ one from assessor M004. Out of the 25 discussion threads, 10 received an overall high score from at least one medically qualified assessor but an overall ‘poor’ score from another, suggesting that doctors do not always agree on the quality of information.

### ***Discussion***

The pilot study suggested that the websites surveyed are capable of producing health information of reasonably high quality. Of 79 assessors, 66 considered the threads they assessed to contain information that was at least somewhat medically/scientifically accurate, and when assessors’ scores were averaged, only one thread (Q21) fell below the middle score band. Four were rated ‘very good’ by the mean of all assessors, 10 ‘good’ and 10 ‘average’.

Question	Respondent	Overall score	Average score	Question	Respondent	Overall score	Average score
Q01	M01	6		Q18	P06	10	
Q01	M02	7		Q18	P05	12	
Q01	P09	9	8	Q18	M04	13	
Q02	M02	7		Q18	P04	15	
Q02	M01	12	9.5	Q18	M05	17	
Q03	P01	10		Q18	M06	18	
Q03	M01	11		Q18	M01	24	15.5
Q03	M02	12		Q19	M05	7	
Q03	P09	13	11.5	Q19	P05	7	
Q04	M02	7		Q19	M04	8	
Q04	P01	11		Q19	M01	9	
Q04	M01	12	10	Q19	P04	15	
Q05	M03	12		Q19	M06	15	
Q05	P02	14		Q19	P06	16	11
Q05	P03	14		Q20	M07	15	
Q05	M01	20	15	Q20	M01	16	
Q06	P02	13		Q20	M05	17	16
Q06	M01	14	13.5	Q21	M04	10	
Q07	M01	7		Q21	P07	12	
Q07	M02	7		Q21	P04	15	
Q07	P01	10	8	Q21	M01	15	
Q08	M01	12	12	Q21	P08	18	
Q09	M04	10		Q21	M05	23	15.5
Q09	M01	11	10.5	Q22	M04	9	
Q10	M04	15		Q22	M05	15	
Q10	M01	21	18	Q22	P07	15	
Q11	M04	6		Q22	M01	16	
Q11	M01	8	7	Q22	P06	16	14
Q12	M01	7		Q23	M08	8	
Q12	M04	7	7	Q23	M04	10	
Q13	M04	12		Q23	P04	13	
Q13	M01	15	13.5	Q23	M01	16	11.75
Q14	M04	11		Q24	M05	11	
Q14	M01	14	12.5	Q24	M04	11	
Q15	M01	13		Q24	M01	12	11
Q15	M04	13	13	Q25	M04	13	
Q16	M01	10		Q25	M05	14	
Q16	M04	10	10	Q25	M01	18	15
Q17	M04	14					
Q17	M01	16	15				

	Very good (5-8)		Average (13-17)		Poor (18-21)
	Good (9-12)				Very poor (22-25)

TABLE 21: *The overall score of each thread as given by each assessor, and the average score of all the assessors for each thread. Some threads (e.g. Q1, Q7 and Q11) were assessed more favourably than others (e.g. Q10, Q21), though there was often a range from good to bad within a single thread (e.g. Q22).*

*N.B. As 1 indicates high quality information and 5 low quality, a lower score indicates higher quality.*

In only two instances did any one of the assessors think someone might make a “very ill-advised” decision based on the information provided: M001 thought this for Q18, and M005 for Q21. In each instance, other medically qualified assessors gave a higher rating. There were only three occasions in which any assessor felt the questioner may be led to act in a way that could put their health at risk (M001 for Q10 and Q18; M005 for Q21). In each case, the question had been assessed by more than one assessor (Q18 by seven people; Q21 by six; Q10 by two) but only one of the assessors felt this way when others did not. In the case of the thread that had received the second lowest rating overall (Q21, on which a parent asked whether it was sensible to expose their five-month-old baby to the child of a friend, who had chickenpox, to “get it out of the way”), comments made in the discussion forum by the original poster, following a series of replies that advised against exposure, suggested the parent had decided against it, yet one respondent, M005, still considered that someone reading the forum would act in way that may put their health at risk. M005 seemed to have missed the fact that while original poster had asked about a particularly harmful course of action, the forum community had strongly advised against it, and appears to have succeeded in steering the poster away from her original plan.

Rating the discussion threads on different criteria enabled me to look more closely than previous studies at how, where, and why poor ratings had been awarded. It is interesting to note, for example, that the controversial discussions around vaccination and herbal/natural remedies in the chickenpox discussions led to 36 separate low ratings of 4 or 5 being awarded by eight different assessors across eight different discussion threads, and just two of those discussion threads (Q18 and Q21) accounted for 56% of all the low scores. These were more often awarded against the inaccuracy of the information (n=11) or incompleteness (n=18) than the information being likely to lead the poster to make an inappropriate or ill-advised decision (n=3).

It was not within the scope of the study to compare those questions that were asked about the more serious conditions (ie, HIV and diabetes) with those that were asked about the milder condition (ie, chickenpox) but I appreciate that these may be important factors in influencing the replies given and they warrant further research. I felt what was more valuable was to look at where the 11 lowest scores were awarded and for what reason.

The worst assessment was given to Q18 by M001, who scored it at 24 out of a possible worst score of 25. The question, on r/parenting, was from a parent who posted a photo of their child's spots and asked, "Is this chickenpox?"

Two doctors (M001, M006) scored the thread poorly, but two others (M004 and M005) considered the quality to be average and the question received an average rating from the mean of all assessors. This shows that doctors do not always agree, and suggests that group wisdom may indeed provide the likelihood of a better answer than reliance on a single expert, as Aitamuro (2016) has suggested. This thread received far fewer votes (a score of +2) than the two other reddit chickenpox threads, which received scores of +63 (Q19) and +59 (Q20), suggesting that the reddit community agreed with the survey respondents' views that information quality was questionable.

While more than half of the assessors (4/7) considered the information given to be 'somewhat' or 'very' scientifically inaccurate and to cover 'very little' of the medical information they would expect to see, six out of the seven respondents did not think this would lead to harmful behaviour. It is also worth noting that some posters on the forum did encourage the original questioner to go to the doctor, who later posted an update to say that they had taken this course of action. This is particularly interesting as it provides evidence that although the information was assessed by some experts to be poor, it did not lead the original poster to dangerous behaviour and suggests that he was able to sort the sensible advice from the mix of replies given.

Qualitative responses given on the questionnaire, and the responses on the forum, pointed to a general attitude that asking for diagnosis over the Internet was a bad idea. Three survey respondents specifically mentioned a long response about herbal remedies which had influenced their judgement, even though this comment had been ignored by the forum (it received no voting activity) and was ordered lower than other, more sensible replies, suggesting that the reddit voting system was signposting the community away from it. This informed my study design by suggesting that in later stages of the study, looking more closely at how this works would be valuable.

The second lowest assessment given overall was by M005 to Q21 (“Chicken pox–is 5 months too young to expose?”). M005 awarded three lowest possible ratings: against the accuracy, appropriateness and usefulness of the information. The survey was completed by six respondents (three doctors, three public) in total and the other five rated the information more favourably (one of the public was critical, but not as critical as the doctor), with the thread receiving a middling rating from the mean of all assessors. The discussion related to a parent’s question about the safety of exposing their five-month-old child to someone who was infected with chickenpox in the hope of getting the disease “out of the way”. The discussion contained a range of views, from some parents who thought there would be little harm in it (largely due to experience of their own children having had the disease at a similar age with no problems arising) to those who considered it dangerous. Several replies actively discouraged the parent from exposing an infant so young. None of the discussions displayed anti-vaccination viewpoints and none actively encouraged the mother to go ahead. At the end of the discussions, the original poster summarized her understanding of the discussions and stated that, after reading the advice, she thought that the ideal age to catch chickenpox “is two to six years,” implying that she had been convinced that deliberately exposing a five-month-old child would not be a good idea. Therefore, it is difficult to

understand why the one assessor (M005) felt that the poster would have made an ill-advised decision that would have put [her child's] health at risk, rather than taking the view of another assessor (M004) who made the qualitative comment: "I think she came to the right conclusion based on the information given." As mumsnet, from which the thread was taken, does not have a voting system, it is not possible to determine how well the online community rated the thread or the individual answers.

Q10 ("I am somewhat prone to depression, but even more so now that I am HIV+. How do you guys deal with it?") about HIV on reddit received two lowest ratings from one of its two respondents (M001), against completeness and appropriateness of information. A qualitative response given in the comments box explained that the low ratings had been given because none of the replies encouraged the poster to seek professional help, which the respondent (a GP) believed they needed. Therefore, it was not so much that poor information was given, but that the appropriate good information was not. Another respondent (M004, a hospital consultant) marked the discussion as average, and it received moderate upvoting (a score of seven) from reddit.

The final low rating was recorded against a discussion on whether diabetes affects a person's ability to recover from a cold ("Does it take longer to get over a cold if you have Type 2 diabetes?"). One of four assessors (M001) felt that the information given was 'very medically/scientifically inaccurate,' but did not feel that it was inappropriate or likely to lead the respondent to act in a way that may put their health at risk.

As only four of the 79 returned surveys were responsible for all 11 instances of low ratings and just one of those (Q18. Reddit/chickenpox) was responsible for five of those 11, the figures would not stand up to statistical analysis, but they do suggest some interesting characteristics:

- Doctors do not always agree with one another.

- The public seem less confident in their assessments than doctors.
- Threads that scored poorly were generally ‘dragged down’ by one bad comment in the middle of a longer discussion, such as the one on herbal remedies in Q18 or the parents who thought there was no harm in exposing the five-month-old baby in Q21.
- Threads that were identified as problematic do appear to have this assessment at least partly reflected in the reddit voting activity.

This suggested to me that there was value in designing the main study survey so that individual comments/replies, rather than entire threads, were assessed by doctors. A thread could contain opposing views and suggest different courses of action across often long discussions; taking a collection of comments and treating them as a single unit of information risked the entire thread being marked down because of one bad answer amongst many.

It is individual posts and comments that are voted on by the reddit community, not entire threads. By focusing on a single post or comment, it is not only possible to determine whether qualified medical professionals consider the information to be high or low quality, but to compare whether the reddit community was able to correctly classify information as good (by upvoting it) or bad (by downvoting) and to order it appropriately. While poor quality information may be found online, if the community recognises it as such and provides signposts of this to other community members, this will lower its message credibility and represents Collective Intelligence, enhancing the health literacy of a single reddit user emerging from the reddit crowd.

The results also suggested that considering how the mean scores awarded by a group of doctors compared with the collective scores of the reddit community may have merit. A direct comparison of ‘very good’, ‘good’ etc may not be appropriate: a better approach would be to see how doctors and the reddit community rated posts in comparison to one another – i.e. out of what was

available, did they agree on what was best, and what was less good, so that the ordering of the posts, if not the absolute scores, would correlate? While such a comparison was not appropriate in the pilot study – only reddit, of the three websites considered indicated how information was received by the online community – it needed to be considered in the context of my research.

I concluded from this that a valuable exercise would be to try to determine the extent to which – if at all – the reddit community upvotes answers that are factually and scientifically accurate, and downvotes answers that are not. If it does, voting is a particularly important technology affordance as the surface credibility and message credibility it provides can guide information seekers towards good information that will be valuable to them during a public health emergency and/or away from information which may be useless or harmful.

#### **6.6 Phase II-B: Investigation of information quality on r/ebola: voting activity and its correlation to doctors' assessments of information**

The votes cast by the reddit community are an important signpost of potential quality on reddit. For most subreddits – including r/ebola – any registered reddit user, regardless of whether they are subscribed to that subreddit, can vote on the content found there<sup>8</sup>. There is considerable literature on voting behaviour on reddit, including whether a small number of active voters is truly representative of the views of all the (more passive) readers (Gilbert, 2013), and acknowledgement that upvoted comments represent those that are popular and most favourably received by participants in online platforms but are not necessarily an indication of quality (Stoddart 2015a, b) but there has been

---

<sup>8</sup> Moderators can set up the subreddit so that only subscribers can vote, an option usually taken by subreddits which attract controversy, such as the political subreddit r/The\_Donald, which supported Donald Trump's successful bid for the U.S. Presidency, though this function only works once one is 'inside' the subreddit, and does not prevent non-subscribers from voting on a post that reaches the Front Page of reddit as a whole.

remarkably little exploration of whether factually accurate information is more likely to be upvoted than less accurate information. A positive correlation between upvoting and accurate information would prove that this feature of reddit can provide a strong signpost to which information and advice users should take on board. A strong correlation between downvoting and poor quality information would also suggest that the downvoting function is a good signpost for which information should be ignored or discarded.

### ***Methodology***

To test the hypothesis that good quality information is likely to be upvoted, and poor quality information likely to be downvoted, I selected 12 questions which were asked on r/ebola between 9<sup>th</sup> September and 28<sup>th</sup> October 2014 at the height of the West Africa Ebola outbreak and the r/ebola subreddit's activity, each of which had attracted several answers on which voting had taken place. These were presented to 27 medically qualified UK professionals (17 members of the Royal Society of Medicine, 10 members of the West Kent Medico Chirurgical Society) each of whom, independently and without sight of the answers given by the others, was asked to rate four of the answers provided to each question as 'good', 'bad' or 'neutral' (using definitions included on the questionnaire – see Appendix IV). The scores they gave were then compared with the upvote-downvote scores awarded by the r/ebola community to see if its members recognised information as 'good', 'neutral' or 'bad' in accordance with medical convention. A more complete explanation of the methodology used for this study was given in Chapter 3.

Though not the primary intention of the study, this approach tests the health literacy of the subreddit community, and thus the agency it might provide for identifying the best answer from several returned in response to a single query. If supported, this would suggest that the voting system acts as an example of

an efficient ordering mechanism needed to help navigate the information available, considered necessary by Lévy for Collective Intelligence and a requirement for the emergence of Crowd Wisdom by Surowiecki.

## ***Results***

Each of the answers taken from r/ebola was rated by each of the doctors independently; they did not have sight of the answers the other doctors gave, nor how the answers had been voted on by the r/ebola community, eliminating any danger of groupthink or bias. These ratings were then translated into a numerical value where good = 3 points, neutral = 2 points and bad = 1 point. An answer which had received a score from all 27 doctors could, therefore, receive scores ranging from 27 points, for a question every respondent considered to be bad and had only awarded one point to, to 81 points for a question every respondent had considered to be good and awarded three points. The scores were then aggregated and three scoring bands were determined for the overall scores: 'Bad' for comments scoring 27-44 overall (coloured red on the data matrix, see Appendix V), 'neutral' for comments scoring 45-62 (coloured blue) and 'good' for comments scoring 63-81 (coloured green); if not all 27 doctors had rated an answer, the scoring bands were adjusted accordingly. If the reddit voting system is indeed an indicator of factual accuracy, we would expect the 'good' band to correlate with answers that were strongly upvoted on r/ebola, the 'neutral band' to correlate with answers which received no strong upvoting or downvoting activity, and the 'bad' band to correlate with answers which were downvoted on r/ebola. This allowed the voting behaviour of the two groups of doctors, and all the doctors together, to be compared with the voting behaviour of the r/ebola community.

The discussion threads were purposely selected to contain comments that had been downvoted and/or removed by moderators, to see if these were treated

similarly by the medically qualified respondents. Therefore, a higher percentage of information in the selected answers is likely to be rated 'bad' than if the answers had been selected randomly. The percentage of 'bad' information found on reddit in this study phase therefore appears to be higher than was found in the pilot study. The percentage ratio of good:bad:neutral assessments in the pilot study is 52:35:13, but here is 23:30:47. This does not indicate that the quality of information on r/ebola was lower than on other subreddits, such as r/diabetes and r/hiv aids, from which posts were taken in the pilot study. It was equally difficult to find posts containing downvoted answers on r/ebola as it was on other health threads and in fact, only nine of the 12 discussion threads selected for the study included downvoted comments while three did not, due to the paucity of choices on offer. The discussion threads chosen for this part of the study were deliberately and purposively those that contained at least some poor information.

The 27 doctors were presented with 12 questions that had been asked in self-posts on r/ebola, and four answers for each of those questions that had been offered and voted on by the reddit community, so 48 answers to rate. The doctors were asked to mark each answer as 'good', 'neutral' or 'bad' and the answers given by the reddit community were also categorised as 'good', 'neutral' or 'bad' according to the voting activity that had taken place on them. An answer which had a score of at least +2, indicating that at least one person had upvoted it, was categorised as 'good'; answers on which no voting appeared to have taken place and which displayed only the default score +1 were categorised as 'neutral'; and answers which displayed a zero or less, indicating that at least one person had downvoted it, were categorised as 'bad'. As the level of voting activity was different on different threads, I chose this method to determine which answers had been endorsed (good), rejected (bad) or neither (neutral) by the reddit crowd, to use as a comparison with the ratings given by the doctors. 'Good' answers from either group are green (G)

on the results matrix (Appendix V), ‘neutral’ answers blue (B), and ‘bad’ answers red (R).

DOC ID	DOC-MEAN OF ALL OTHER DOCS	SIG LEVEL	DOC-REDDIT	SIG LEVEL	DOC-ONE DOC
	A	B	C	D	E
RSM-M01	0.18	0.12	0.33	0.02*	0.12
RSM-M02	0.73	1.00E-05*	0.58	1.40E-04*	0.44
RSM-M03	0.52	1.10E-04*	0.43	0.002*	0.32
RSM-M04	0.35	0.01*	0.24	0.05	0.22
RSM-M05	0.27	0.03*	0.19	0.12	0.17
RSM-M07	0.28	0.03*	0.23	0.07*	0.18
RSM-M09	0.67	1.00E-05*	0.42	0.002*	0.42
RSM-M12	0.67	1.00E-05*	0.60	1.00E-05*	0.41
RSM-M13	0.60	1.00E-05*	0.39	0.0032*	0.37
RSM-M14	0.79	1.00E-05*	0.51	1.60E-04*	0.48
RSM-M15	0.64	1.00E-05*	0.60	2.00E-05*	0.40
RSM-M16	0.55	4.00E-05*	0.44	3.20E-04*	0.34
RSM-M17	0.49	3.40E-04*	0.48	5.20E-04*	0.31
WKCS-001	0.36	0.01*	0.19	0.12	0.22
WKCS-002	0.60	1.00E-05*	0.37	0.005*	0.37
WKCS-003	0.69	1.00E-05*	0.37	0.005*	0.42
WCKS-004	0.73	1.00E-05*	0.45	0.002*	0.48
WCKS-005	0.63	1.00E-05*	0.53	1.70E-04*	0.39
WCKS-006	0.52	1.50E-04*	0.46	0.001*	0.33
WCKS-007	0.64	1.00E-05*	0.35	0.007*	0.40
WKCS-008	0.78	1.00E-05*	0.42	0.002*	0.47
WCKS-009	0.73	1.00E-05*	0.52	8.30E-04*	0.46
WCKS-010	0.65	1.00E-05*	0.42	0.001*	0.39

TABLE 22: *Correlations of each doctor’s scores to the average of all other doctors (column A) and to the reddit RBG scores (Column C), along with their significance levels for a one tailed test on the null hypothesis of zero correlation. The far-right column shows the mean of the correlations of each doctor with each of the other doctors (column E).*

Correlations on the scores given by doctors who answered more than half of the survey questions were then run for each doctor against the mean score given by all the doctors; the scores given by reddit; and the score given by each other doctor. Standard Pearson correlations are used throughout and statistical significance is assessed using Monte Carlo randomisation on 100,000 random reorderings of the answers. Results are shown in TABLE 22.

Column A – Doc-all docs: This shows the correlation of each doctor's scores with the average of all the other doctors' scores (taking into account missing values). All but one of the p values is  $<0.05$  (marked \* in Column B).

Column B – Sig level: This shows the significance level of values in Column A, with regard to 100,000 correlations taken with randomised re-orderings.

Column C – Doc-Reddit: This shows the correlation of each doctor's scores with the Reddit RBG scores (taking missing values into account). The p values are shown and all but three are less than  $<0.05$  (marked \* in Column D).

Column D – Sig level: This column shows the significance level of the values in Column C with regard to 100,000 randomised re-orderings. (One way only – i.e. the probability that a random reordering of the scores would produce a correlation as high or higher).

Column E – Doc-one doc: This column shows, for each doctor, the mean correlation of his/her scores with answers of the other doctors.

A scatterplot (FIG: 39, below) visualises how each doctor's scores correlate a little better with the average score of all other doctors than they do with the RBG score of reddit. The correlation between the reddit RBG and the average score of all doctors is 0.67 ( $p < 10^{-5}$ ).

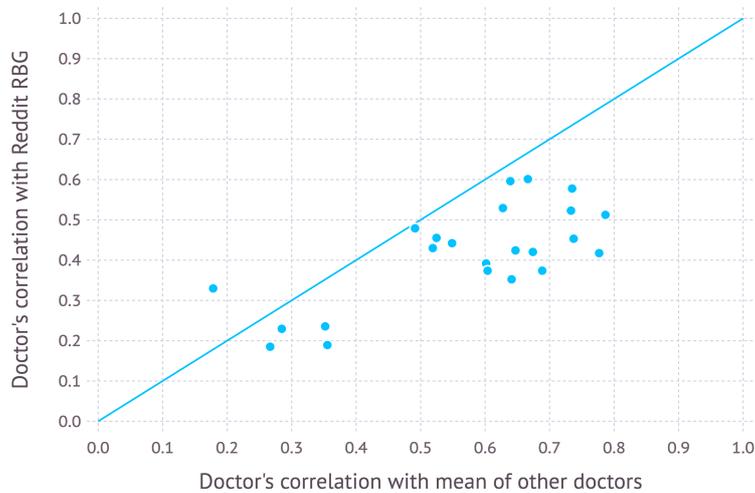


FIG 39: *Correlation between each doctor's score with the mean of other doctors and each doctor's correlation with the reddit RBG score.*

It is interesting to note that those doctors whose scores do not correlate highly with the reddit RBG scores (shown in the low cluster group) also have a low correlation with the average score of all other doctors. This suggests that those who disagreed with the reddit community were also likely to disagree with their medical colleagues.

Of most interest is that the average correlation of each doctor to each of the other doctors is 0.57 – lower than the correlation to the average of all doctors. In most cases, the correlation of reddit with the average of all doctors – 0.67 – is better than the correlation of a single doctor to the average of all doctors, suggesting that reddit is closer to the crowd of doctors than all but five of the doctors. Typically, one doctor's opinions are more likely to resemble reddit than they are the opinions of another doctor, although this finding is not statistically significant. In terms of crowd wisdom, this not only suggests that consulting a crowd of doctors will provide a higher quality result than consulting a single doctor, but that if a crowd of doctors is not available,

consulting the reddit crowd may still be more accurate than consulting a single doctor. As actual health consultations rarely take place with more than one individual doctor, this finding warrants further research.

There were four instances in which the answer scored most highly by the r/ebola community was different to the answer most preferred by at least one of the doctor groups. These are as follows:

- (QD.d-9.2): the r/ebola community strongly upvoted an answer which told a user asking about the safety of travelling to work on a bus in Cleveland, USA (where no cases of Ebola were present) that he was, “totally paranoid and need[ed] to calm down”. Both groups of doctors disliked this answer, preferring one that patiently explained he was in virtually no danger. While blunt replies pointing out there was no reason to panic tended to be upvoted, they did spark a debate in the private modmail forum about whether the community was being too harsh on users who were genuinely concerned, even where concern was misplaced, and if it should try to foster a more sympathetic atmosphere to such questions. The reddit community did upvote the sympathetic answer the doctors preferred, just not by as much.
- (QE.a-8.3) The r/ebola community preferred an answer explaining that the reason for the current Ebola outbreak being larger than past ones was more due to the political and economic situation in the region than any characteristic of the disease, whereas the doctor groups both preferred an answer which essentially gave the same information but was worded slightly differently. However, neither group rated any of the answers provided as ‘bad’. The answer preferred by both doctor groups displayed a reddit score of +1,

suggesting that no voting activity on it had been recorded. It may have simply been missed by the reddit community.

- (QI.d-4.2) The r/ebola community particularly liked an answer which used an analogy of a crashing plane to explain why Ebola seemed to be treated as more dangerous than SARS when it was also being described as less infectious. The WKCS group particularly disliked this analogy, though the RSM group was ambivalent to it. The two doctor groups disagreed on what the best answer actually was, the RSM group preferring one which explained the difference between the medical terms ‘infectious’ and ‘contagious’ and the WKCS group preferring one which explained the difference in terms of case fatality rates. The doctor groups disagreed on which answer they preferred – and both their preferred choices were upvoted by the r/ebola community, just not as strongly as the aircraft analogy answer.
- (QL.d-1.2) The r/ebola community most liked an answer which attempted to explain, in jargon-free terms, why healthcare workers were becoming infected. The respondent explained how difficult it was to maintain good hygiene precautions in challenging conditions. Both doctor groups preferred answer QL.a-1.3, which explained the same concept in a slightly different wording. Neither group voted any of the answers as ‘bad’ however, and there was little to separate the scoring on any of the answers across either of the groups.

### ***Discussion***

Overall, there was broad consensus between the groups on what constituted ‘good’ answers – for 8 out of 12 questions reddit agreed with at least one of the doctor groups on which answer was best, and there was only one occasion

on which the r/ebola community's most preferred answer was voted 'bad' by both doctor groups – the dismissal of the concerned bus commuter.

The doctor group was, overall, less likely to rate answers as 'good' than the r/ebola community were to upvote answers (only 12 out of the 48 answers – 25% – received a 'good' score from at least one of the doctor groups and only 4 (8%) received a good score from the combined doctor groups, whereas 27 out of 48 (56%) were actively upvoted by the r/ebola community). This is consistent with the findings from the pilot study which suggest that while information on reddit is of reasonable quality, it is perhaps not necessarily perfect nor the best information available anywhere. Nonetheless, if other options are constrained, the findings do suggest that the voting system is more likely than not to direct its users to the better answers of those available – the maximized choice – which may be particularly important where access to more professional healthcare advice is constrained. The upvoted answers on r/ebola do appear likely to represent the best of what is available from the reddit crowd and to correlate well with the assessments made by doctors.

Out of the 48 answers provided across the 12 questions, 41 received both 'good' and 'bad' scores from the individual doctors surveyed, and for three out of the 12 answer sets the two doctor groups disagreed collectively on which was the best answer of the four provided for that question, suggesting (as in the pilot study) that there was not always consensus amongst medically qualified individuals on what was indeed a 'good' answer. Of the 19 answers considered 'bad' by the crowd of doctors, seven received a 'good' rating from at least one doctor, and all four of the answers considered 'good' by the doctor crowd received a 'bad' rating from at least one doctor. This suggests that if only one doctor is consulted, it is reasonably likely that their opinion may deviate from the crowd; therefore it follows that consulting a crowd of doctors may have advantages, just as consulting a crowd of laypeople does.

	RSM	WKCS	Docs	Red
QA.a-12.3	#	#	#	#
QA.b-12.4	*		*	
QA.c-12.1		*		*
QA.d-12.2				
QB.a-11.3	#			#
QB.b-11.4		#	#	
QB.c-11.1	*	*	*	
QB.d-11.2		*		*
QC.a-10.3	#	#	#	#
QC.b-10.4		*		
QC.c-10.1				
QC.d-10.2	*		*	*
QD.a-9.3				
QD.b-9.4	#	#	#	#
QD.c-9.1	*	*	*	
QD.d-9.2				*
QE.a-8.3	#	#	#	*
QE.b-8.4				=#
QE.c-8.1	*	*	*	=#
QE.d-8.2				
QF.a-7.3				=#
QF.b-7.4	#	#	#	=#
QF.c-7.1				
QF.d-7.2	*	*	*	*

	RSM	WKCS	Docs	Red
QG.a-6.3		#	#	=#
QG.b-6.4	#			=#
QG.c-6.1	*	*	*	*
QG.d-6.4				
QH.a-5.3		#	#	#
QH.b-5.4	#			
QH.c-5.1				
QH.d-5.2	*	*	*	*
QI.a-4.3		*	*	#
QI.b-4.4	*			
QI.c-4.1				
QI.d-4.2	#	#	#	*
QJ.a-3.3	#	#	#	#
QJ.b-3.4				
QJ.c-3.1				
QJ.d-3.2	*	*	*	*
QK.a-2.3	#	#	#	=#
QK.b-2.4				=#
QK.c-2.1				
QK.d-2.2	*	*	*	*
QL.a-1.3	*	*	*	
QL.b-1.4				#
QL.c-1.1	#	#	#	
QL.d-1.2				*

Table 23: Table showing voting behaviour on the r/ebola subreddit in relation to answers offered to self-post questions regarding Ebola.

	Red denotes answers downvoted by reddit and awarded low scores by doctors
	Blue denotes answers neither upvoted nor downvoted by reddit; middling scores from doctors
	Green denotes answers which were upvoted by reddit and received high scores from doctors
#	# indicates lowest scoring answer given by a particular group
*	* indicates the highest scoring answer given by a particular group

With regards to the answers that were considered ‘bad’, and therefore required signposting that would steer users away from them, the three groups – of RSM doctors, WKCS doctors and the r/ebola community – agreed on 6 out of 12 (50%) occasions on which was the worst or equal worst answer<sup>9</sup> overall of those provided. On a further three occasions the two doctor groups disagreed on which was the worst answer but one of the doctor groups agreed with the r/ebola community, giving a total of 9 out of 12 (75%) occasions on which at least one doctor group and reddit agreed on the worst answer. In two of the three occasions where there was disagreement between reddit and either of the doctor groups on which was the worst answer (QL/1 and QE/8) none of the three groups considered any of the four answers on offer to be bad; reddit upvoted answers to which the doctors gave the lowest score but these were considered neutral rather than bad overall by the doctors, so unlikely to cause actual harm, though it does suggest the reddit community missed spotting the best answer of those available.

Only in one case – (QI/4) did the reddit community choose as its favourite an answer the doctors considered to be bad (QI.d-4.2) and even then only one of the doctor groups (WKCS; RSM voted it neutral). The answer preferred by r/ebola is a non-medical anecdote attempting to explain risk using an analogy to a crashing airplane, which would not have resulted in the questioner coming to harm (and I have to say, as a professional working in risk management, I found the anecdote clear and useful, and struggle to understand why the medical community disliked it; I am not sure the answer they preferred actually answered the questioner’s query, rather than pointing out that he was misunderstanding certain medical terms).

---

<sup>9</sup> For some questions, one answer had been downvoted and another removed. As removal may result from offensive or racist comments as well as factual inaccuracy, it is difficult to tell which answer would have received the most downvotes had the removed answer been allowed to remain. In these instances, a heavily downvoted answer and a removed answer are treated as ‘equal worst’ in the opinion of the reddit community.

The RSM group rated 17 of the 48 answers bad, and 19 were given this rating by the WKCS group. Of these, 12 of the 17 scored 'bad' by the RSM group were also downvoted and/or removed by the r/ebola community (71%). A further three received only the default +1 vote, suggesting that they were ignored (or may have been missed) by the r/ebola community. Only two (4%) were upvoted by the reddit community: (QD.d-9.2), the overly harsh attitude to the concerned commuter discussed above, and (H.c-5.1), a flippant answer to a question about whether or not Ebola can survive on paper notes and coins which did nothing to answer the question but could not cause harm.

Of the 19 answers scored 'bad' by the WKCS group, which marked slightly more harshly overall, 11 were downvoted and/or removed by the r/ebola community (58%), and further three received only the default +1 vote, suggesting that they were ignored, or may have been missed. In total, five answers rated 'bad' by the WKCS group were upvoted by the r/ebola community, but of these, one (QC.d-10.2) was also considered the best of the answers offered by the RSM group; two related to the harsh response to the commuter and the flippant answer regarding money, discussed above; and the remaining two were both in relation to answers given to QI-4, both of which the RSM group had given a 'neutral' vote to overall. Overall, there were five out of 48 occasions on which the reddit community had upvoted an answer this group would have considered bad (10%). When the doctors' scores were considered together, the reddit community upvoted three answers (QD.d-9.2, QH.c-5.1 and HI.c-4.1) a total of 6%, considered to be bad by the doctors.

The only occasion on which the r/ebola community upvoted an answer considered to be 'bad' overall by both sets of doctors and which could conceivably lead to harm was the case of (QD.d-9.2), the overly harsh attitude to the concerned commuter: ridicule may result in a health seeker withdrawing from the information source and missing further information. As previously

discussed, the r/ebola moderators themselves showed concern about whether such treatment of worried users was fair. It is also worth noting that while the answer received a poor overall score from both doctor groups, four doctors surveyed rated it as ‘good’, and four rated it ‘neutral’, again suggesting that there was no overall medical consensus. One who rated the harsh reply ‘bad’ added the qualitative comment, “... but he does have a point”.

As mentioned above, 41 of the 48 answers which received a ‘bad’ rating from at least one doctor received a ‘good’ rating from at least one other. The seven remaining questions received only ‘bad’ or ‘neutral’ answers – there was no answer which received only all ‘neutral’ or all ‘good’ answers from both doctor groups (though QA.b-1 did from the RSM doctors while QE.c-8.1, QF.d-7.2, QH.d-5.2 and QJ.d-3.2 did from the WKCS doctors).

Of these seven, five were downvoted and/or removed by the r/ebola community, one received no apparent voting activity, suggesting that it was ignored (or missed), and the final one, which received minor upvoting was a flippant answer that made no attempt to answer the question and stated an opinion that could not lead a health-information seeker to come to harm.

QUESTION	DOCTORS COMBINED	REDDIT
QF.a-7.3		Downvoted (-16pts)
QF.b-7.4		Removed by moderators
QG.a-6.3		Removed by moderators
QH.a-5.3		Removed by moderators
QH.b-5.4		No voting activity (+1pt)
QH.c-5.1		Minor upvoting (+2pts)
QJ.a-3.3		Downvoted (-2pts)

TABLE 24: *Treatment by the r/ebola community of answers to which none of the doctors surveyed gave a ‘good’ rating.*

A correlation of 0.67 ( $p < 10^{-5}$ ) between the average score of all doctors and the RGB scores of the r/ebola community suggests that Collective Intelligence in the form of health literacy can emerge from an online health forum. A user with little health literacy of their own can have their agency increased by using voting scores as a signpost on the forum to guide them towards the best, or at least better, information and away from the lower quality information. r/ebola would seem to be slightly better at guiding users away from the worst of the information than providing or guiding them towards the best, but as a mechanism for maximising choice, the hypothesis that upvoting behaviour on r/ebola correlates with better quality information and downvoting behaviour correlates with poorer quality information has been upheld.

It is interesting to note that there was not always consensus over what constituted 'good' and 'bad' answers amongst the doctors. None of the 48 answers received consistent scoring from all 27 doctors surveyed, and 41 out of 48 answers received scores of both 'bad' and 'good' – though as in the pilot study, the doctors were more likely to agree with one another (and with the r/ebola community) on what was bad information than on what was good information. This lack of consensus amongst the medical professionals suggests, however, that consulting a crowd may be better than consulting an individual even where a single expert is available, particularly where the 'expert' group – in this case doctors – lacks direct experience of the specific context under discussion (in this case an Ebola outbreak).

The scores given by the medical professionals and the r/ebola community have another interesting characteristic in that while the r/ebola community could see how others voted on the forum, the doctors voted in isolation from one another and without sight of the voting that had taken place on the forum. The overall agreement suggests that the r/ebola forum was not influenced by groupthink and that votes cast on answers was a fair reflection of the genuine

opinions of the individual voters, unbiased by the views of others around them. This suggests that a forum such as r/ebola can meet Surowiecki's criteria of the clever mechanism required to harness the Wisdom of Crowds.

### **6.8 Phase III: Observation of reddit and interviews with moderators**

The process of selecting the discussion threads for use in the Phase II-A study and the question and answer sets for Phase II-B highlighted that examples of bad information online are hard to find. This includes examples of what, according to my understanding, appeared to be incorrect and also of posts and comments that had been heavily downvoted. Examples were particularly hard to find in the chronic condition subreddits. The two studies also showed that what little bad information there is tends to be called out by the subreddit users in a display of Collective Intelligence by the community. Information that is obviously wrong or misleading neither appears nor remains in health forums with any regularity, often because the forum moderators remove such posts. In the Phase II-B study, four answers that the doctors scored low had in fact been removed from r/ebola, and I had only been able to access them because of moderator privileges I had been offered. This suggested that the moderators play an important role in regulating the quality of the information on the forum. I therefore felt it would be useful to interview reddit moderators about their experiences of moderating health forums, and of the quality of information they have experienced while doing so. The answers they gave revealed a new dimension to how the quality of information is maintained on reddit and on individual subreddits.

#### ***Interviews with moderators***

In total, I interviewed 14 moderators of health-related subreddits to ask them about the type of information that was posted on their forums and how they managed poor quality information. Full methodology is covered in Chapter 3.

Before discussing the moderators' impressions of the quality of information found in their subreddits, it is worth examining their attitudes to conventional science and medicine, and to doctors. None of the moderators I interviewed gave the impression that they saw either themselves or their forums as a proxy for a doctor, and none expressed negative or antagonistic opinions of doctors or of the medical profession in general.

Some expressed disappointment that they could not always discuss issues in as much depth with their doctors as they would like – in particular, one [RM006] felt that doctors did not always fully consider the negative impact unpleasant side effects could have on patients, and were unwilling to discuss the option of not taking medication in such circumstances, whereas the community could help with this – but in general, those interviewed respected and deferred to the medical profession and to medical and scientific convention while also recognizing that some support functions were better provided by the peer community, consistent with the existing literature (Wadley et al, 2014). A typical response was: “Some medical information ought to come from an authoritative source, ideally from your own doctor who knows your specific circumstances. Other medical information is best found from other patients, things like the day-to-day details of how to cope, what to look for, what to expect” [RM001], another said “what drug should I take – that’s a question for your doctor” [RM004], and “90% of the time ... we will try to discourage people from saying say, use this [prescription drug], rather than, talk to your doctor about this” [RM006].

The moderators liked to use peer-reviewed scientific papers where possible to support the information presented on their subreddits. As one said, “I think people [on reddit] respect scientists and engineers and people who are deeply rooted in science ... if there’s an interesting scientific study that shows up, then we’ll try to add that, kind of talk about it where possible” [RM002]. They

also considered official health organisations to be trusted sources of quality information: “say a paper got posted from the [U.S.] National Institute of Health, well, that’s a really reliable source. I’m not going to doubt what it’s saying. We actually have scientists and doctors and very important smart people on reddit,” pointed out a chronic condition moderator [RM005], “you know, real doctors and pharmacists who are posting really great articles to do with their job, or submitting their own papers that have been cleared for medical journals”. “How do you check or verify [information’s] correctness? Comparison to established sources”, [RM007] was a common response to how moderators decided whether information was accurate.

This respect for doctors and scientists was also apparent in the r/zika and r/ebola strands. The moderators valued, appreciated and encouraged the involvement of experts, mentioning the advantages of having, “medical staff, doctors, nurses, scientists, sharing information with each other, people in the field sharing notes” [RM010] and of having, “contributors who are doctors, including at least two who actively worked in outbreak areas.” [RM007].

This was not only true of the outbreak subreddits: a moderator of r/science reported valuing Ebola information from, “.... graduate students and professors who work in infectious disease and epidemiology. We gathered as many experts as we could, and for a while we even had an Ebola researcher [...] We also had several prominent professors in epidemiology and infectious disease come and talk directly to the reddit group” [RM018].

From the interviews conducted, it seems safe to assume that the moderators would agree with my definition of good quality information: that it should be in accordance with current medical policy and consistent with the information that would be provided by a qualified health professional.

### *Characteristics of poor quality information*

The interviews suggested that there are different types of bad information found in online health discussion forums, with some more prevalent in certain types of forum than others. In general, the moderators of the chronic condition subreddits I spoke to had experienced very little information that could be considered medically inaccurate or which they thought could possibly lead to actual harm. “I haven’t seen anything quite that bad” [RM004] was a typical response when the question was asked directly. When pushed, one moderator guessed they may have encountered such examples “perhaps once or twice a year” [RM001], while RM006 replied “I can’t remember any incidences of factually incorrect or misleading information on the site. Not off the top of my head”. On the rare occasions when the subreddits did encounter information that was thought to be wrong or potentially dangerous, moderators reported that this was quickly corrected by themselves or by the community: “Bad info is almost always corrected, or at least debated, by the other users. We are fortunate to have a good feeling of community here and people are generally protective of one another.” [RM001]. One challenge, reported by RM003, was that posters could sometimes fail to recognize, or actively ignore, that there may not be a one-size-fits-all answer in health. “[Each person’s] condition can be very different, and it varies a lot from individual to individual. What’s right for one person may not be right for another. We do have to temper statements and say this worked for you, in some cases it might be helpful but in a lot of cases it might not”.

Moderators also reported attempts by vendors of various therapies – many of dubious efficacy – to sell their products to people with a known health condition who could be desperate to get better. “Be careful, because there’s a lot of people out there who will take advantage of you and your money in your

quest to get better”, warned RM001. Posts attempting to sell something were generally removed immediately.

The most common type of ‘bad’ information reported by the chronic condition moderators related to herbal remedies and alternative treatments, which could perhaps more accurately be categorised as biased rather than (or as well as) inaccurate. Moderators felt that posters tended to exaggerate the likely benefits, rather than that there was no value in them at all. They, and other community members, would generally try to inform the original poster of what is and is not supported by medical evidence, and provide links to scientific papers and further information on official sources such as the CDC website, while also allowing posts to remain, particularly those that provided personal anecdotes of alternative remedies that had provided some relief: “...I do think that the supplements and treatments and pharmaceuticals and even some of the pseudoscience stuff, cumulatively if it helps the person’s quality of life then that might help to manage [some of] the symptoms, then I think that will be considered a success from a practical medicine sense as well as a quality of life sense. We try to support things that are able to do some good and are limited on harm”, reported RM001.

A standard tactic used by the moderators to debunk and debate information considered to be incorrect, incomplete and/or biased was to refer the original poster to peer-reviewed scientific papers and official sources of information such as the WHO and CDC, via comments made on the post. “I don’t usually delete stuff”, explained RM004. “I see things like get your Himalayan Salt Wraps and I post pretty conclusive and scientific information about how this is not actually valid treatment,” while RM002 reported that when there is discussion and debate in the subreddit, she will often, “go to r/medicine or r/pharmacy or r/biotech [...] and say, ‘hey guys, I moderate this subreddit, I wondered if any of you have expertise in this particular article or subject we’re

talking about, could you come in and talk to us and give us your professional opinion on the matter’, which is nice, because once you have someone who proves they’re an authority in their field, people tend to just shut up and listen.” This points to another aspect of the Collective Intelligence of reddit – the crowd of each subreddit can expand when needed by reaching into others, to which the site provides unique access and a shared sense of community. It also suggests that experience and knowledge of reddit’s structure, beyond the individual subreddit and the topic to which it is dedicated, enables the moderator to look for such support. This may play an important part in the quality control process, and warrants further consideration.

In the experience of the moderators I spoke to, and my own observation of the r/ebola and r/zika subs, the outbreak subreddits seemed less prone to posts regarding alternate remedies or non-conventional treatments than the chronic condition ones. This may be down to the fact that the community was Far-at-Risk and not actually affected by the disease to the point that treatments were needed; the WAG interviewees reported several instances of alternate remedies and quack cures being peddled to the Near- and Real-at-Risk communities in West Africa. Posts about specific treatments may pose a challenge for an outbreak forum if cases of the disease did become widespread within the posting community, particularly as while the moderators of the chronic condition health forums are generally ‘expert patients’ (Wilson, 2001) who have lived with their conditions for a long time and have had extensive contact with both the medical profession and others living with the condition, the outbreak subreddit moderators had less direct experience. Of the seven chronic condition subreddit moderators I interviewed, all but one – who worked for a charity supporting those living with the condition – had the condition themselves. The health literacy of the chronic condition moderators with regard to their specific condition is generally high and this puts them in a good position to help guide and inform other members of their community,

particularly newer ones. They were generally familiar with a range of academic and scientific literature published on the condition itself (and of where to find it), and with scientific literature which had tried and failed to find any proven benefits for common alternative therapies linked to the condition or its symptoms. This could help to counter any exaggerated claims for those therapies. They were also adept at checking source characteristics to help determine quality: they would regularly check the value of new reports or papers posted to the forums by looking for previous papers which had been published by the author, and assessing where they had been published. This requires experience of the condition and of reddit's signposting structure.

The moderators of the disease outbreak subreddits are not expert patients, but they may not need to be. Factually inaccurate and biased information reported by r/ebola and r/zika fell more into the categories of conspiracy theories and sensationalised media stories. These were uncommon on the chronic condition subreddits – as the conditions tend to have years of scientific and medical research behind them on causes/correlations – but were a problem for the outbreak subreddits. While r/ebola did not see the level of conspiracy theories that spread across West Africa at the height of the outbreak and which were also reported by the NGO interviewees in the Phase I study – including conspiracies that black magic could offer protection from Ebola; that Ebola was being spread in certain areas of Sierra Leone by the government to kill off its opposition supporters; or that Ebola was being greatly exaggerated (or, conversely, ignored) by the governments of poor West African countries to leverage Western aid – it did experience posts suggesting that the virus was likely to, or had, become, airborne. The r/zika subreddit received several posts relating to the conspiracy theory that genetically modified mosquitos were linked to the microcephaly and birth defects that were also being linked to the Zika virus. The moderators reported frustration that such stories – which were constantly refuted by the scientific community – were often upvoted.

↑ 7  
↓

**Speculative** Ebola airborne: A nightmare that could happen [CNN] cnn.com  
submitted 1 year ago by [aquarain](#)  
23 comments share save hide give gold spam remove approve report nsfw flair

all 23 comments  
sorted by: **best** ▼ set as suggested sort enable contest mode make announcement

[-] [CurlSagan](#) 12 points 1 year ago\*  
 ↓ The MSF doctor who did an AMA yesterday answered a question about this, saying:  
 | No reason why Ebola should mutate to an airborne virus. It survives in body fluids, but not on dry surfaces  
 | due to it's lipid membrane.  
 |  
 | However, I take the fears of epidemiologists and virologists about an airborne ebola to be the sexy, press-  
 | friendly shorthand version of their worries about the virus mutating in general with the result of changed  
 | transmissibility, lethality, or symptoms.  
 |  
 | Since it seems like everyone in the English speaking world doesn't give a shit, hyperbole couldn't hurt.  
 |  
 | permalink embed save spam remove give gold

[-] [briangles](#) 6 points 1 year ago  
 ↓ The MSF doctor said there was no reason for it to go airborne. That does not mean it can't or won't. The  
 | doctod was not a virologist and didn't talk about about selective preasures. It's not impossible nor is it  
 | impossibly unlikely. It is less likely than not that it will go airborne. But it's better to prepare for this than to  
 | sit on our hands shouting into the echo chamber that we're all going to be fine.  
 |  
 | The exponential growth rate means (if the WHO numbers are right, and I highly doubt that) will be 10,000  
 | by September 31- October 1 (or in actuality 20-30,000) That sort of growth allows for an insane amount of  
 | mutations. I read a report of 6 doctors who were infected while not working. There is a minute chance that  
 | there is a mixed strain of airborne in some of these cases already, but it was transmitted via fluid.  
 |  
 | Well know for sure in a few months.

FIG 40: A post implying the Ebola virus could go airborne was upvoted. The most highly voted comment refutes the story, but many redditors may only read the post title and not necessarily click through to the comments.

↑ 2  
↓

**Blog** Pandora's box: How GM mosquitos could have caused Brazil's microcephaly disaster ✓ theecologist.org  
submitted 6 months ago by [SouthMoon](#)  
9 comments share save hide give gold spam remove report nsfw flair

all 9 comments  
sorted by: **best** ▼ set as suggested sort enable contest mode make announcement

[-] [SouthMoon](#) [S] 1 point 6 months ago  
 ↓ idk why this does not show, it would be nice if you can do something about it? Is citing articles from long  
 | before this happened, their assertions can be tested and what they said is very interesting.  
 |  
 | permalink embed save give gold approve [ removed by AutoModerator (remove not spam) ]

[-] [IIWIIMS](#) 1 point 6 months ago  
 ↓ The answer may lie in a sequence of 'jumping DNA' used to engineer the virus's mosquito vector - and  
 | released into the wild four years ago in the precise area of Brazil where the microcephaly crisis is most  
 | acute.  
 |  
 | It cites speculative information. Something not needed at this point. It is a conspiratorial posting steeped in  
 | FUD (fear, uncertainty and doubt). When the 'assertions' have been tested and proven, please feel free to  
 | post that information.  
 |  
 | permalink embed save parent spam remove give gold

[-] [SouthMoon](#) [S] 0 points 6 months ago  
 ↓ It makes non sense what you saying, because you are posting information about the microcephaly,  
 | which has yet to be proveen is because of zika. Also the article, cites information from long before all  
 | this mess started. From 2001, 2009, and 2014  
 |  
 | permalink embed save parent give gold approve [ removed by AutoModerator (remove not spam) ]

FIG 41: A post on r/zika claiming there may be a link between the cases of microcephaly in Northern Brazil and genetically modified mosquitos. This is challenged in the comments, some of which have been removed.

The promotion of alternate remedies on the chronic condition subreddits and prevalence of conspiracy theories on the outbreak subreddits is consistent with the existing literature: that online health information is more prone to being biased or incomplete than it is to being downright inaccurate (Impicciatore et al, 1997; Eysenbach et al, 2002; Whitelaw et al, 2014). In both cases, the discussion forum format enabled community members (and the moderators) to provide additional information to correct, complete or balance the discussions, for example the r/zika moderators were quick to point out that papers pointing to a possible link between genetically modified mosquitos and birth defects in Brazil were authored by Dr Mae Wan Ho, a well-known anti-GMO crops conspiracy theorist who was attracting much attention on r/conspiracy, a subreddit known to support and propagate various anti-establishment theories.

This again points to a value in moderators having experience of reddit and its processes beyond experience of the topic to which their forum is dedicated. The way in which chronic condition moderators and the outbreak moderators reported that they verify information quality on behalf of the community suggests sophisticated use of the surface credibility and message credibility markers discussed earlier in this chapter. Information quality is easier to verify in link-posts than self-posts, which moderators report are more likely to contain the poorest quality information. “Information in posts are typically easier to confirm based on the source, if the source is broadly accepted as being trusted, as is the case with WHO, MSF and CDC,” explained RM008. The exception was link posts to blogs, which were generally considered to be common sources of poor quality information. They are heavily discouraged and often outright banned by most of the health subreddits (and, in fact, are frowned upon across reddit in general, regardless of the subject focus of the subreddit). “If it’s a random person’s blog then those are pretty likely to be bad” [RM013] was the typical attitude, with another remarking that blogs are, “invariably of low quality” [RM001]. Moderators routinely remove any posts

linking to personal blogs and the rules of most subreddits state this clearly. Self-posts often contain opinions and statements which are not backed up by evidence – and so are also generally considered to be less reliable.

This again points to the value of moderators understanding how reddit is structured and organised, and how their experience of the positioning of other subreddits, and of the credibility signposts of posts and posters, might play an important part in the quality control process.

### ***Differences between posts and comments***

Moderators reported that lower quality information is more likely to be posted in a reply to an original post than to be an original post itself. “If bad information is posted, generally it’s as a response to someone else’s post, it’s someone commenting on their post, giving them advice [...],” reported RM003. This does, however, provide an easy way to manage such information: moderators and other forum users can post a second comment, correcting the first, or providing a more evidence-based response. “Sometimes people will make claims that they heard from somebody who heard from somebody and it’s not good information. We can say actually, that’s not correct, or it’s not correct in all circumstances, and here’s more information about it, if you’re interested,” RM003 explained.

The greater prevalence of problematic content on the disease outbreak subreddits compared with the chronic condition subreddits came not so much from fundamental differences between the two types of forum, but because such content more often comes from new posters who know little about the topic, and who may be misinformed – perhaps taking sensationalised information they have seen in the media at face value. Also, chronic condition subreddits do not tend to experience the sudden, rapid increase in traffic

experienced by the outbreak subreddits and which leads to large numbers of new, poorly informed users coming into the forum at once. Highly speculative and misleading media stories are less likely to be published on topics which have a long history of scientific and medical research behind them, leading to less poor quality information outside of reddit that new or existing users can link to.

How to deal with poor quality information presents a challenge to the moderators. Removing information can lead to accusations that they are stifling free speech and censoring users, particularly where scientific information is scarce or issues are still unclear and rife for speculation, but leaving such posts *in situ* to be (hopefully) challenged by the community leaves open the possibility that incorrect information may be upvoted, with the quality and reputation of the subreddit suffering as a result. An issue for r/ebola and r/zika was not only the quality of information, but the quantity of low quality information that coincided with the sudden emergence of the disease outbreaks and a corresponding large influx of non-expert users to the forums, especially as this came when the situation was fast-moving and facts were sometimes uncertain. Even information from legitimate sources could sometimes turn out to be wrong, or was misinterpreted by the media, such as high predictions of possible total cases of Ebola by the CDC. Many media sources picked up only on the top end value of modelling predictions (up to 1.4 million cases – but only in a scenario in which no infection control was practiced allowing the virus to spread freely) and reported the story in way that implied the CDC was saying this *would* happen *no matter what*, not that it *might* happen *if no action was taken to prevent it*. Without the full context, this could undermine confidence in official sources, an effect that has parallels with the increased influence of international news media – the so-called ‘CNN Effect’ (Robinson, 1999) – in situations where government policy is seen to be poorly articulated, unclear or lacking (Robinson, 2000). A challenge for

moderators in such circumstances is to balance the community's desire for information with information of sufficiently high quality to be of value.

### ***The moderator as an administrator and technician***

Beyond describing the types of bad information experienced on the forums, the process of interviewing the moderators revealed something I had not anticipated at the start of the study, and which provides a fascinating insight into how reddit operates.

A large amount of what the subreddit moderators do is not only related to correcting, completing or countering information posted on their forums: they also carry out a huge amount of 'back office' maintenance and management of the forums. This is virtually invisible to the users but is essential not only to running a platform, but to running a *good* platform. Understanding this may help in the setting up and operation of a public health forum during a future disease outbreak but there is little coverage of it in the existing literature.

The first issue that highlighted this was that moderators reported often needing to deal with troublesome and inappropriate behaviour from subreddit visitors and contributors, ranging from the purveyors of dubious medical 'cures' and therapies trying to take advantage of people with long-term health conditions, described above, to general Internet trolling and racism. r/ebola experienced, "forays into the subreddit by the delightful racist residents of /r/GreatApes and /r/coontown, who were revelling in the outbreak. The trolls, the preppers<sup>10</sup>, the worried wells, the conspiracy theorists... It could feel like King Canute, ordering back the tide" explained RM008, an r/ebola moderator, when asked

---

<sup>10</sup> 'Preppers' is a term used to describe people who believe Western Society is headed for a breakdown which will necessitate a return to a less modernised, more feudal society, for which they should prepare by learning survivalist skills and hoarding food and other essentials that will be needed when the breakdown comes.

to describe experiences of poor quality information. In this situation, the usually reliable voting system can become a disadvantage, as groups of newcomers deliberately pushing certain narratives can downvote posts and comments that run contrary to their position while upvoting those that support them. The r/ebola moderators felt that an ‘Ebola is airborne’ conspiracy was fostered by one particular user who had a reputation for causing trouble across reddit, remarking, “there’s no denying the tone on occasions became noxious. Fear does not always bring out the greatest attributes in people! We had people wishing Kaci Hickox [a nurse who refused to comply with quarantine arrangements] caught Ebola, or demanding she was executed for treason, or wanting Amber Vinson [a nurse who was subsequently diagnosed with Ebola] charged with attempted murder for flying on a plane” [RM010].

Spam was another problem, and one also experienced more frequently by the outbreak subreddits than the chronic condition subreddits, as spammers tend to target subreddits that are attracting increased attention, with posts appearing on the Front Page. “[Rapidly growing] subs are the hunting ground for spam artists,” explained RM008. “They will make [a] posting completely unrelated to the sub or its threads”. Removing the postings that came with this could take up significant time, and again suggested that there was value in a more detailed examination of not only the role of the moderator, but of the processes and tools the moderator uses to perform their role, many of which are not apparent to the forum users. This informed the final stage of my study design, which will be explained and covered in Chapter 7.

## **6.9 Conclusions**

The quantitative research carried out in Phase II-A (the pilot study) and Phase II-B (assessments of the information on r/ebola) suggests that when a health information seeker poses a question to the reddit crowd, the answers they get

back are more likely than not to be of reasonable quality. The signposting available on the site tends to flag up poor quality information and steer users away from it, as do other forum users. This suggests that reddit does have technology affordances that help users to identify the better information and make the maximized choice out of those available to them. Characteristics of reddit – and of individual subreddits – including surface and message credibility markers, and the voting system, do help users to navigate the information it collects to find the better items and locate potential experts who can answer additional questions. This enables Collective Intelligence to emerge and provides a level of health literacy beyond that of the individual user, increasingly their agency accordingly.

This chapter has, therefore, answered three of the subresearch questions:

- SRQ-2, *What technology affordances of online discussion forums might help users to trust the information found there?*  
Characteristics such as flair, the indication of the source and the visibility of voting behaviour all provide surface credibility, signposting information the community considers to be of better quality information. These guide users towards better information and away from poorer quality information. The organisation of information into subreddits with clear positioning, where experts on the topic gather, also affords trust.
- SRQ-3: *Is the quality of information in online health discussions of sufficiently high quality to be of value to health information seekers?*  
The answer is broadly yes, with a 0.67 correlation between doctors and the reddit crowd on what information is consider to be good.
- SRQ-4: *What characteristics of discussion forums help to maintain or compromise information quality?* The findings presented here suggest

the answer to is closely tied to the actions of moderators, but there is far more to the role of a subreddit moderator than just the ability to recognise good information and ‘call out’ the bad. The process of moderation is virtually invisible to the average reddit user but it has a huge impact on how the forum can be run and maintained effectively.

Exactly how moderators undertake their role and how this affects the quality of the information found in online health discussion forums will be explored at length in the following chapter.



## 7: WHO ARE THE MODERATORS AND WHAT DO THEY DO?

---

### 7.1 Introduction

During Phase II-B of the study, it became apparent that the poorest quality health information on discussion forums is often removed by the forum moderators, who play an important role in filtering out bad information and ensuring that what remains is factually correct and appropriate. What was most interesting, however, was not only what the moderators said about what they do, but also about *how* they do it.

The discussions I had with 18 forum moderators, and my observation of the moderation logs of r/ebola, r/zika and r/science, suggest that moderation has a significant technical component, such as spam removal, setting spam filters, configuring style sheets and designing graphics. There appears to be a plethora of moderation activity that goes on behind the scenes of which the average user of reddit is largely unaware. To undertake their role effectively, the moderator needs three distinct skill sets:

- **subject matter skills**, comprising topic expertise, which influences the content of the subreddit. This can include, in the case of health subreddits, experience of living with the condition as well as medical and scientific knowledge about it.
- **technical skills**, requiring some basic programming ability. This determines how the subreddit functions and what use the moderator can make of the tools and technical functionality available to them.
- **context skills**, which relate to the moderator's experience of the moderation process and of reddit, rather than of the topic to which their subreddit is dedicated. This experience influences how quickly they

can respond to challenges and how easily they can locate help to boost their existing skills when needed.

There appears to be little awareness of this need for different skills in the existing literature, particularly of the need for technical and context skills. Previous academic interest in moderators, of reddit and of online forums in general, has tended to focus on the social aspects of the role such as how they welcome new members to the forum and help them to settle in (Hsieh et al, 2013); how they demonstrate norms and expected behaviour of the community (Lampe and Johnston, 2005); and how they respond to negative behaviour exhibited by other users (Duggan 2014, Choi et al 2015). These studies recognise the need for subject matter skills, and some for context skills related to subject matter issues, but do not look at the technical skills needed to keep the forum operational, how this affects the moderation process nor how moderators learn and develop them. A recent study by Park et al (2016) has examined how technology can help *The New York Times* to select the highest quality comments made on its news stories, identifying a shift in perception from moderation being only a ‘policing’ role to being a more valued editorial role, but this focusses only on how the technology sorts the comments, not on the relationship and interactions between the human moderators, the moderation tools available to them and the technical skills they need to make the most of those tools. This represents a significant gap in current knowledge. In fact, Golder and Donath deliberately excluded moderators from their 2004 study of ‘personalities’ and roles observed in the online communities of Usenet – such as lurker (a passive user), troll (a deliberately disruptive and/or abusive user) and newbie (inexperienced newcomer) – on the basis that ‘moderator’ was a technologically imposed role, rather than one that emerges naturally. This misses the point that if technical skills are required, this may limit who can be a moderator or how effective the moderator can be, which in turn may also influence the efficacy of the forum.

The reason technical skills might have been overlooked is that the need for them is virtually invisible to anyone but the moderators performing the tasks for which they are necessary. The studies cited above only observed activity on the forums but did not interview the moderators or users; in consequence, the authors may not have been aware that such processes were going on.

This highlights a further gap in the literature: as the need for technical skills is under-reported, so too is how moderators learn the skills they need to moderate effectively. There is some recognition that experience of subject matter is important: Shultz and Nakamoto (2013) have stressed that in health, ‘procedural knowledge’ of why certain procedures are followed, enabling the patient to make a reasoned choice about the proper course of action, can be as important as the ‘declarative knowledge’ of medical facts and figures. Several studies show that experienced posters on peer-to-peer networks help to present information in jargon-free language that is easier to understand (Ancker and Kaufman, 2007; Ancker et al, 2009; Huh et al, 2013), keep conversations on-topic (Kassing, 2015), guide users to good sources of information and advise them to seek professional help when necessary (Coulson and Shaw, 2013), though moderators may also tend to ‘push’ users off the forum towards professional medical help too quickly (Huh, 2015). It has also been identified that reddit health forum moderators tend to fit the definition of ‘expert patients’, with experience of the condition being discussed on their forum. Moderators of r/stopsmoking, for example, tend to be long-term quitters who can offer advice and support to those at the beginning of a journey they have already made (Wadley et al, 2014). These studies focus on the moderators’ experience of the health condition, however, not on their experience of the forum, its norms, or the tools used to configure its settings. Nor do they consider the wider context in which such discussion forums sit – for instance, does the platform only host one forum, or is it one amongst many, and how does this affect what is available to the moderator?

### *Technology Affordances for Moderation*

The reddit platform provides its moderators with considerable support. This is not only technical support in the form of an existing platform onto which they can build their forum, but also support from other moderators and experienced posters across the wide reddit crowd, both inside their own subreddit and from others. Most subreddits have a team of moderators rather than a single moderator – but no studies, to the best of my knowledge, have so far looked at the processes, advantages and disadvantages of moderation by a small group of formal moderators who have system privileges most of the users on the forum do not. The dynamics of this and its influence on information provision has been underexplored. Lampe and Resnick (2004) and Lampe and Johnston (2005) have studied ‘distributed moderation’ by a team of moderators (in large discussions on Slashdot.com, a site with a similar structure to Reddit covering a number of computer and technology-related topics) concluding that the voting behaviour of ‘moderators’ helps to show new users what is and is not acceptable posting behaviour. Slashdot ‘moderators’ are a small percentage of its volunteer users, however, who are temporarily afforded voting privileges, while the real content management is undertaken by the site’s professional editors, guided by the moderators’ votes. Its ‘moderators’ are not the day-to-day managers and administrators of the forums in the same way as reddit’s are, nor do they need to work collaboratively to perform their moderation duties.

These issues are important to consider as they affect and influence reddit’s Collective Intelligence. On the one hand, the Collective Intelligence of the moderator team might lead to a better process of moderation than could be achieved by one moderator alone, particularly if the group displays the diversity of problem-solving abilities identified as conducive to CI by Woolley et al (2015). Then there is the Collective Intelligence of reddit as a whole – the knowledge stored collectively within the community – on which

the moderator team, or any member of it, can draw. Buntain and Golbeck (2014) found that Reddit users rarely participate in more than one community or a tightly-knit circle of communities, with few participating in subreddits on disparate topics – for example a user who participates in a subreddit for a specific rare health condition may also participate in r/rarediseases, or in both r/football and the subreddit for a particular football club – yet all but one of the reddit moderators I interviewed moderated more than one forum (range 2-28, falling into two groups who either moderated a small set of 2-8 or a larger set of 20-28, with both sets containing subreddits on diverse topics). As reddit users who are interested in multiple topics play important roles in large and viral conversations, often helping posts to achieve the critical mass needed to reach the Front Page (Choi et al, 2015), this deserves more attention but there has been little in-depth examination of this group activity or of whether it plays a specific role in the moderation process.

I will use this chapter to set out why the three skill sets of subject matter skills, technical skills and context skills are required, and how the reddit system is well-configured to support moderators in gaining them through the Collective Intelligence of moderator teams and of the full reddit crowd. To do this, I will first define a reddit moderator, then show what roles moderators take on, and finally explain how reddit provides the technical and context support needed.

## **7.2 Definition of a reddit moderator used in the study**

In the context of this study, I define ‘moderators’ as individuals who are formally recognised by the reddit platform as moderators of one or more subreddit. These people have certain system privileges afforded to them resulting from that status that are not available to all reddit users. To meet this definition, the reddit user must appear in a list of moderators on one of the subreddit forum front pages (see FIG 42 below).



FIG 42: Moderators of a subreddit are listed on its front page. The box shown here lists the moderators of the r/ebola forum as of 3 Jan 2017, including myself. Four of these (plus a fifth moderator who subsequently left the forum) were interviewed for this study.

Note that the list in FIG 42 also includes moderation tools that can be set up by the human moderators: these include Botwatchman, software that removes posts from, and bans, suspected bots reported to reddit by human moderators and users, and Modmail\_archivist, which automatically moves moderator mail to a designated private subreddit. This is the first indication that there is a technical, as well as human, aspect to moderation.

The box also shows that (on this forum at least) moderation is undertaken by a team of moderators, rather than a single individual. I will return to who these individuals are, how they become moderators and how moderation teams are formed, later in this chapter.

### ***Where moderators sit in the hierarchy of reddit***

All moderators, even those who moderate a forum alone, are part of the wider reddit crowd, and it is therefore important to understand where they sit within this. Reddit's Collective Intelligence Genome has some elements of *Hierarchy*, which largely relate to increasingly privileged system access as one moves up that hierarchy.

**Administrators:** At the top of the hierarchy are the Administrators, the paid employees of reddit. They set site-wide rules and make site-wide system modifications. They can ban users and remove moderator privileges. In extreme circumstances, they can closedown a subreddit or force it to go private, meaning that it is not visible to the public. Within a subreddit, however, the Admins have no specific influence. Any votes they cast, or posts they make, have no greater weight than of any other user and they cannot make modifications to the subreddit settings: they can only affect default settings across the whole of reddit.

**Lead Moderator:** The Lead Moderator (my term – there is no official term used by reddit) is the reddit user who set up the subreddit and has access to its full suite of moderator tools and system privileges. Their name is top of the list of moderators shown on the subreddit. The Lead Moderator can at any time appoint new moderators to join the team, and remove existing ones; they can confer approved poster status and grant flair to other users (see Chapter 4); and can configure the subreddit settings. To set up a subreddit, and thus to be a Lead Moderator, the reddit user needs to have built up some experience of participating on reddit, and to have earned 100 karma points.

**Moderator:** Once the subreddit is operational, the Lead Moderator can choose to appoint additional moderators, who may be given full access to the suite of moderator privileges and tools, or only some of them, depending on the preferences of the moderator who appoints them. A new moderator may, for example, be given system privileges that enable them to approve or remove comments flagged as needing moderator attention, but may not be able to change subreddit settings. Some very large subreddits, such as r/science, have ‘layers’ of moderators with an according hierarchy of privileges: an ‘inner circle’ has access to all moderator privileges, while ‘comment moderators’ can do little more than remove comments or approve/remove reported comments.

Moderators move up through the layers based on demonstration of appropriate action. A user can be a moderator of more than one subreddit at any one time.

Seniority of moderators is determined by the order in which they join the moderation team. Once an additional moderator is added, they may then appoint others. Moderators with full system privileges can remove any moderator appointed after them, but not ones who were appointed before.

**Approved posters:** ‘Approved poster’ is a status that can be conferred on a registered reddit user by a moderator and is given to people who are considered to make consistently useful and high quality contributions. Approved posters cannot make any modifications to the system, but their status means that the posts and comments they contribute skip any holding queue the moderators may have imposed, and appear immediately on the site.

**Subscribers:** Reddit users can choose to subscribe to a subreddit they are particularly interested in. In the case of public forums, this adds no system privileges except that posts from the subreddit may appear on the user’s personalised Front Page. Users must subscribe to Restricted forums to be able to post, comment or vote, however, and for Private forums, only subscribers can even read the forum. In the case of Restricted and Private forums, subscribers will usually require an invitation to subscribe from the forum moderator(s). Subreddits can be configured so that only subscribers can vote.

**Registered users:** Users can register with reddit through a function on the homepage, creating a username and profile. This enables them to subscribe to subreddits of interest, and to submit content to reddit in the form of posts and comments. Registered users can also vote on content submitted by others. An individual is at liberty to register more than once with different usernames, which they may use to post in different forums, or for posts they make in a single forum. A moderator could choose to post in their own subreddit using a

different account from the one tied to their moderator privileges, for example, if they did not want their status to be obvious to other posters.

**Unregistered users:** Unregistered users can read content that appears on the Front Page and on all public subreddits, but cannot post, comment or vote.

The reddit *Hierarchy* is unique in that while the Admins sit formally at the top, all active content management goes on only at the subreddit level. Content is contributed and voted on by registered users, and managed by the volunteer moderators, who have considerable system privileges on their subreddits (though not across the whole of reddit), but do not own the subreddit(s) they moderate. The moderators can influence the look of the forum, and determine what remains on it and what is removed (and what is prevented from reaching the subreddit in the first place, through the spam filters and configuration settings). For the content they do not remove, the community collectively *Decides* by *Voting* what appears most prominently. The votes cast by moderators or Admins have no greater weight or influence than the votes cast by any other redditor. Like ‘ordinary’ users, moderators can only vote once: a second vote from the same account cancels out the first. Posts and comment order, and what reaches the Front Page, is determined by voting activity alone: it cannot be influenced by the moderators or the Admins.

The moderators, who outnumber the administrators considerably (by approximately 500:1<sup>11</sup>), are essential for the day-to-day running of the subreddits even though they are unpaid volunteers. They receive little material reward for their efforts past the opportunity for another reddit user to reward them with ‘reddit gold’ (a credit payment which covers a limited-time subscription to premier services) but they are the hierarchical level that has the

---

<sup>11</sup> There are approximately 26,000 moderators on reddit, and approximately 50 Admins. Figures are estimated from discussions with the moderators of larger forums and questions asked on reddit forums r/modhelp and r/theoryofreddit.

greatest influence on information quality, and it is difficult to see how reddit could operate in the way it does without them, or how the current business model could support the platform if the moderators had to be paid. With a few notable exceptions, the Admins rarely interfere with what the moderators do (Weninger et al, 2013). A useful analogy of the roles is to see the Admins as the publishers, and the moderators as the editors, with reddit analogous to a publishing group producing information on many different topics including politics, DIY, sport, humour and health. The Admins are not moderators but their influence in the moderation process should not be ignored. I will consider it where appropriate throughout this chapter.

A registered user, a moderator and an Admin – i.e. any member of the *Crowd* – can contribute information to the reddit *Collection* of information. Each has equal agency to post a link to information elsewhere on the web or to submit a self-post. It is also the reddit *Crowd* collectively – via the *Decide by Voting* of the hotscore (see Chapter 4) – that determines what is considered good, with each individual member of the crowd able to vote only once. As one moves up the *Hierarchy*, additional privileges confer no greater influence on contributing good information. Also, if the *Crowd* collectively decides to upvote controversial material, there is very little that those higher up the *Heirarchy* can do to prioritise information they consider to be better. Additional privileges instead increase the ability to remove or prevent bad information: registered users can only downvote or report bad information, but moderators can remove it and ban the users and domains that are likely to be its sources. Admins can ban troublesome users from the whole of reddit and, in extreme cases, closedown a ‘bad’ subreddit entirely.

Ordering of information on reddit is a *Collaboration* between the *Crowd*, but the *Hierarchy* can ensure that the *Crowd* has only better quality information to choose from. How the moderators enact this will be considered in the following sections.

	<b>INFLUENCE POOR QUAL INFO</b>	<b>INFLUENCE HIGH QUAL INFO</b>
<b>Admins</b> - system-wide access	Closedown subreddits Force offensive subreddits to go private Remove a lead moderator	Handover an abandoned/poorly moderated sub to a new moderator
<b>Moderators</b> - access to moderation tools and system privileges on their subreddit	Counter claims in posts and comments Remove content Ban posters Ban domains Flair low quality sources Set rules Downvote	Start subreddits Designate approved posters Assign other moderators Post good quality information Comment on other posts Upvote
<b>Approved posters</b> - some system privileges	Counter claims in posts and comments Report bad information Downvote	Post good information immediately Comment favourably Upvote good information
<b>Registered users</b>	Make correcting/countering comments (may be delayed) Report bad information Downvote	Post good information (may be delayed) Comment favourably Upvote good information

TABLE 25: *How the different hierarchical levels of reddit influence the quality of information on the site.*

### 7.3 Moderation roles

In the following section, I will cover the roles described by the reddit moderators during the interview process and relate these to the influence they have on the quality of information found on the forum. The full methodology for selecting and interviewing the moderators was explained in Chapter 3.

The categorisation of roles are mine, based on the language and descriptions used by the reddit moderators interviewed and are therefore subjective. They are intended to give an overview of the roles required of the moderators and the skills needed to execute them efficiently.

### *Quality Controller*

Given the focus of the Phase II-A and II-B studies on information quality, it is not surprising that there was extensive discussion of the quality control aspect of the moderation role in the interviews – the semi-structured questionnaire I used guided moderators towards it and this may be the reason why all 18 moderators specifically refer to it. Comments that the role of the moderator is, “For me personally [...] about submitting good quality articles”, [RM014] and, “Getting accurate and true information to the reader” [RM005] were typical. Moderators posted and commented extensively themselves to help ensure information they considered to be high quality reached their forum. Most had regularly contributed to the forum itself or to reddit before they became moderators: the route into moderation for 15 of those interviewed was as a regular poster about the topic to which their forum was dedicated, often on the forum itself. The remaining two [RM004 and RM009] had also been regular posters on reddit, but not on the specific topic to which their subreddit was dedicated before setting it up (the other was the non-reddit moderator I interviewed, whose experience of reddit was not relevant here).

Discussion of the moderator’s role in weeding out bad information as well as in shepherding good was common: “you tend to get some posters who put some of that kind of pseudo-science stuff and as moderators you need to steer them away without censoring, necessarily [RM002]”, and “balancing the conversation to ensure there aren’t biased or non-evidence based facts being thrown around and making sure people are getting the correct, most accurate information to make the best choices for themselves” [RM004]. This recognises the need for procedural as well as declarative knowledge (Scultz and Nakamoto, 2011, 2013) which the moderators, as experienced posters, were well-placed to provide.

This quality control role often saw moderators checking the source of link posts to check where the information had originally appeared. “Many of our posts were links to [personal blogs] and they were invariably of low quality. They were the sort of things that nobody would ever have thought to share unless they had written them themselves” reported [RM001] while another saw their role as, “Mak[ing] sure all links posted are accurate. No false information” [RM009], which included flaring posts from media sources, official sources such as CDC and WHO, or peer reviewed journals. This quality control role required not only subject matter skills, but also context skills that enabled the moderator to know that reddit provides these surface credibility indicators to the original source of the information and to the poster. They also need some technical skills to be able to use the tools provided by the platform to flair the original source for others’ benefit.

As well as managing the quality of the content *per se*, moderators also spoke of the need to act quickly on poor quality information: “moderating posts and comments can carry a sense of urgency ... if falsehood is allowed to remain, readers may make the assumption it is endorsed. Once it takes hold and comments start rolling in, it becomes very difficult to counter and correct” [RM008]. Subreddits therefore need to be moderated as much as possible in real time, potentially in different time zones to ensure 24-hour coverage, which may require a large and/or carefully selected moderation team. Some had deliberately addressed this and sought out co-moderators who could provide round-the-clock cover [RM004; RM005], others had recognised the need for it but found that new moderators from other time zones came naturally when sufficient numbers of users from those timezones were active.

Moderators considered quality control to be an important role but also one that was helped considerably by the forum community of posters and voters, who would bring information to the moderators’ notice where required (any user

can ‘report’ a post or comment via a button beneath it, or send moderators a personal message of complaint), display their disapproval by downvoting or disagree/correct the information in a comment. The size of the subreddit crowd, and its ability to contribute to the quality control role in this way eases the burden on the moderator. Quality controller tends to be a collective role between the moderator(s) and the user community, rather than one undertaken by the moderator alone.

### ***Policing the forum***

Another commonly recognised, and widely studied aspect of the moderator’s role is that of policeman. Eight of the moderators mentioned active policing of the forum as part of their role, with three aspects needing to be ‘policed’: poor quality information, offensive behaviour, and spam. Policing poor quality information required subject matter skills; offensive behaviour could require some context skills as well as technical skills; and spam removal required technical skills to configure the forum’s spam filters. One moderator described their role as, “[b]asically keeping control of the content, not necessarily censoring or anything like that, but just making sure that there’s no spam and the conversation stays on topic” [RM011]. Others referred to “Getting rid of hate speech etc [RM009], “to stop people posting that kind of horrible diatribe” [RM015] and to “defend [community members] against unnecessarily hostile comments” [RM001]. Keeping control of discussions and countering antisocial behaviour seemed to go hand in hand, although “don’t silence people’s opinions” [RM009] was also important. Context skills could be important in identifying repeat offenses: one moderator [RM005] spoke of suspected drug addicts coming onto the forum to get information on medical conditions so they could fake symptoms to a doctor and get access to prescription opiate-based drugs. After a while, moderators and experienced

forum users would recognise the traits but newer users may not be able to, and the policing role became particularly important in such situations.

Most recognised the need for this contextual understanding, even if they didn't see it as their primary role: "If you're not removing certain things then [...] people start going into discussions about things like politics, fighting about religion ... it can get very, very nasty and very toxic very, very quickly if someone doesn't step in" [RM015] but the prevailing attitude was, "wherever possible, I let the group police itself" [RM001]. Moderators preferred to be light touch than dictatorial (apart from r/science, which tightly controls its content, allowing peer-reviewed links only). Except for racism and other hate speech, which was not tolerated on any of the forums I examined, moderators preferred the softer approach discussed next to a hard 'policing' role.

### ***Guide/Protector***

Rather than seeing themselves as police officers of the forum, most moderators (n=14) spoke more in terms of being a guide/protector, particularly on the chronic condition subreddits. An important part of their role is protecting users of the subreddit who have lower health literacy, less expert knowledge or who are less familiar with the platform than themselves. This could be considered part of the Quality Controller role, but there is a more social aspect to it, which relies on experience of the condition and contextual awareness of the forum, rather than just medical and scientific knowledge. "We don't want people wasting their time and money and energy on stuff that isn't going to work for them", said RM005 in regard to removing links to dubious alternate therapies that were regularly posted on the forum: "fruity, hippy dippy bullshit not peer-reviewed, crap that can get people sicker than they already are. I see [the role of moderator] as a gatekeeper, somebody to

keep them safe because a lot of people prey on them [offering] magical cures that aren't really cures”.

Moderators also saw themselves as having a role in, “telling people when to get help” [RM006], particularly in terms of when symptoms or behaviour they were describing sounded like something that should be discussed with a medical practitioner rather than on the forum – though Huh (2015) has suggested that moderators may be too quick to discourage users from asking for diagnoses from their peers and that this can be a disadvantage of the moderation process.

An interesting aspect of the Guide/Protector role, for which context understanding was particularly important, was warning potential new moderators of the downsides of moderation. Regular posters on chronic condition subreddits often succumbed to their condition, and for moderators, this often did equate to losing a friend. In the case of the outbreak subreddits, “r/ebola had a very dark side and it wasn't for everybody ... There was no vaccine, no cure, people were becoming infected at an alarming rate and a good number of them were dying. [People volunteering to moderate] took a second look at the situation and decided they were not interested” [RM008]. The Guide/Protector role appeared to sit somewhere between hard policing and the softer Socialiser role discussed below, requiring some contextual experience beyond just subject matter knowledge to know what challenges the forum was likely to experience, from which the community might need protecting and guiding away.

### ***Socialiser***

Only seven of the moderators expressed attitudes towards their role in line with the previously observed role of ‘volunteer socialisers’ (Hsieh et al, 2013)

who help to make newcomers feel welcome. Five of these were chronic condition subreddit moderators; a sixth referred to her role in a subreddit other than r/ebola, about which she was being interviewed. A typical example of this is, “[I see the role of the moderator as] welcoming newcomers and answering all the basic questions everyone seems to have ... Newcomers often feel isolated, overwhelmed and they are often poorly informed... I want their first experience here to be a good one, I want them to feel there are others here who will understand and be able to offer sensible advice. I like to be the first to reply to a new user, to welcome them, patiently share what I think they need to hear” [RM001]. Others saw their role as, “keeping it to a useful community that if someone walks into, they think, oh, this is something I want to join” [RM004], or as “building and maintaining a community ... to build places where people can come to discuss what they want to” [RM015].

This was more common within the chronic disease subreddits, though one r/ebola moderator [RM008] reported: “we wanted people to visit, read and understand”. In the disease outbreak forums, however, moderation was more about providing quality information than a welcoming place to meet and interact with others.

These descriptions of quality controller, police officer and socialiser are consistent with the existing literature on moderators and moderator roles.

The way the moderators described their roles suggested they wanted to steer, rather than lead discussions. It was not of paramount importance that they were themselves subject experts: subject matter expertise could come from the posters, as had been the case in the Quality Controller role. On r/ebola, for example, the moderators reached out to subject experts on r/infectiousdiseases and on the discussion forums of another platform, ProMed Mail, to provide the expertise they felt they lacked.

What the moderator does need to perform their role well, however, is a basic level of technical skill that enables them to set some functions to run automatically in order to reduce the potentially heavy time burden of moderation. This requirement, which has previously been given little consideration, is best described in comments that can be grouped into two role categorisations: *Managerial/administration* and *Engineer/technical*.

### ***Manager/Administrator***

Fifteen of the moderators reported that their role involved day-to-day work that is best described as carrying out mundane ‘housekeeping tasks’ – necessary, but more akin to chores than anything else. “...To clean up the thread, to make sure people aren’t being too inflammatory ... approve the different posts ... check if there’s stuff that doesn’t sound reputable...” explained RM002, while RM008 spoke of, “Posting new information, reading comments posted, updating stats, working through the daily mod[eration] queue, vetting posts for accuracy, responding to misconceptions.... flair, rules, style sheet and subreddit settings ... the daily maintenance and upkeep required”. The reading out of the tasks itself suggests they did not garner high enthusiasm, but were still seen as a necessary part of a moderator’s role.

There was some crossover between this role and that of Quality Controller, but more in terms of the quantity of information than quality. “Keeping down the signal-to-noise ratio and keeping discussions relatively helpful” explained RM004, “to keep discussions on topic” said RM013. This reported level of administrative burden was upheld by the data available from the moderator logs. For example, of 1438 moderator actions recorded in the moderation logs of r/zika between 28<sup>th</sup> January and 1<sup>st</sup> April 2016, 59% were administrative, 37% concerned with content management and just 4% related to policing, such as banning posters.

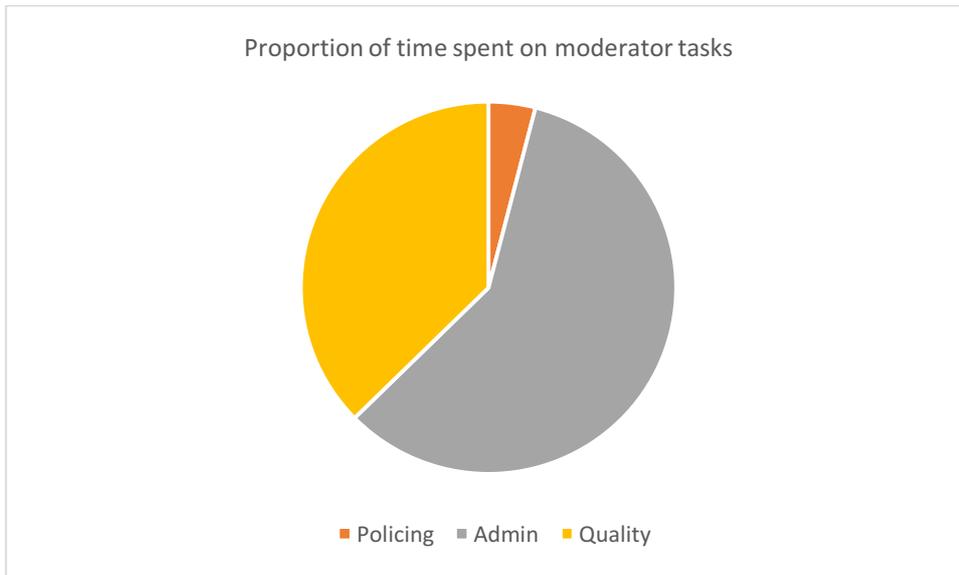


FIG 43: *Proportion of moderators' time spent on Policing, Administrative and Quality Control tasks, recorded by r/zika mod logs, 28<sup>th</sup> Jan-1<sup>st</sup> April 2016.*

The ability to perform this moderator task effectively was at least as dependent on a moderator's technical skills as their subject matter expertise and forum experience, as described below.

### ***Engineer/Technician***

The most surprising finding from the interviews with the moderators, and one that I feel is a unique contribution to knowledge in the field of discussion forum moderation is that the technical skills and effort required to maintain a subreddit are substantial, not least because the technical skill of the moderator can have a strong influence on how time consuming or difficult other roles, such as management/administration and quality control of the content, will be. Twelve of the moderators made direct reference to technical skills they, or other members of their moderation team, needed to have to perform their role effectively, some of which are basic programming skills, some of which

require the user to learn systems specific to the reddit platform, which can only be done ‘on the job’ once the role of moderator has been accepted.

A good technical moderator can set up several automatic features to filter out content that would otherwise have to be weeded out by hand, thus improving message credibility, and making the subreddit more visually appealing, increasing its surface credibility. This was described as, “making the subreddit easy to use for people and making the design appropriate to that subreddit ... getting traffic and basically organising. Filling out [the wiki page], keeping it as a central resource for people” [RM011].

Technical roles could be permanent, or a temporary offer of help to a subreddit needing a quick fix: “I see the crisis happening and I come in and set Automoderator, and set it up for them”, explained RM015, who is experienced at dealing with racist abuse. The use of such features can be crucial to ensuring that the moderator’s limited time is not eaten up in administrative functions, leaving more space for the more visibly valuable socialising, protecting, quality control and managerial skills.

Exactly how the processes of technical moderation operate and influence the quality of health discussion forums has received little attention, however.

The reddit moderators I interviewed were generally aware of the different aspects of their moderation roles: “Moderating has two separate sides: vetting the information in posts and comments and handling background operations” described RM008. RM007 stated that there are, “Three things, really – to maintain the standards of the sub (eg. deleting misinformation, abuse and spam), to lead discussion (by sharing informative and relevant links) and to handle administrative tasks (such as updating the sidebar and header, clearing posts from the spam filter and communicating with and banning users)”.

How the three skills sets I have identified relate to the roles described above are summarised in TABLE 26.

SKILLS REQUIRED	MODERATOR ROLES AND DUTIES
<b>Subject matter skills</b>	Recognising good content and bad content and treating appropriately; submitting appropriate content to the site; flairing sources from which content is likely to be good; knowing when to refer posters on to more professional help <i>Needed for: Quality controller; Protector-Guide; Policing (content); Socialising</i>
<b>Technical skills</b>	Configuring the look of the subreddit, updating the wiki (which determines what appears in the side boxes, etc), adding graphics, configuring Automoderator, setting spam filters, banning posters and domains, configuring flair, managing mod queue <i>Needed for: Quality Controller; Policing; Protector-Guide; Administrator; Socialiser</i>
<b>Context skills</b>	Setting the rules and norms of the forum both formally and informally through interaction; banning members who violate site rules; Recognising likely challenges; Awareness of technical tools available and how to use them; experience of where to find other moderators and reddit users with skills needed for the forum context. <i>Needed for Socialising; Protecting; Administrator; Policing (behaviour)</i>

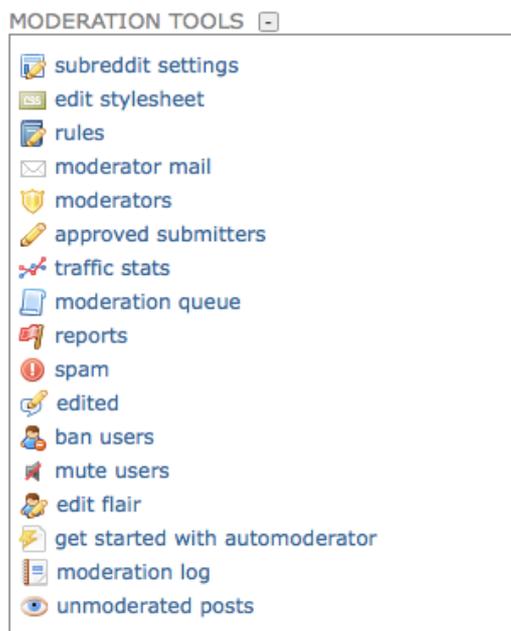
TABLE 26: *Skills needed by reddit moderators*

Interestingly, most moderators reported only becoming aware of the importance of the technical skills and context knowledge after they had taken on the moderator role, and this could prove challenging. The one non-reddit moderator I interviewed, the professional moderator of a forum on HealthUnlocked, had significant technical skills as well as subject matter

expertise and experience of forums, and referenced the importance of all three in his interview, so this does not seem to be unique to reddit.

In the following section, I will explain what technical tools are available to the reddit moderator; how those tools can be best used to aid the process of moderation; how the moderator requires context skills in order to know that the tools are available to them; and how the process of becoming a moderator helps to build this context awareness with the support of the existing moderators and the Collective Intelligence of the reddit Crowd.

#### 7.4 Technical skills: using the moderation tools available



The moderation process on reddit is not only undertaken by human moderators. There is a suite of moderation tools, provided *de facto* by the reddit system and accessed through a box on the subreddit homepage, though this box is only visible to, and accessible by the moderator(s). The tools enable modifications to be made to the site and help the moderation process. The full suite is shown in FIG 44 and explained in TABLE 27.

FIG 44: *The moderation tools box appears on the subreddit homepage of users with moderator privileges to that subreddit. It is not visible to other users, who may be entirely unaware that such tools exist.*

TABLE 27: *Moderation tools and functions available to subreddit moderators*

MODERATOR TOOL ICON	FUNCTIONS ENABLED
 <b>subreddit settings</b>	<p>This tool allows the moderator to configure subreddit settings, including populating the permanent text boxes on its front page; set if it is public, restricted or private; define what types of links are allowed; and decide on the strength of the spam filter.</p>
 <b>edit stylesheet</b>	<p>This allows the moderator to change the appearance of the front page, by adding images or changing the font, background colours and so on. Some knowledge of CSS is needed to use it.</p>
 <b>rules</b>	<p>This is where the moderator can set the rules for what is acceptable or not on the subreddit. These usually appear as a list on its front page. Violation of them can be used as a justification for removing a post or banning a user.</p>
 <b>moderator mail</b>	<p>Allows the moderators to see any messages that have been sent to the moderation team via the 'message the moderators' button on the subreddit, and to reply collectively.</p>
 <b>moderators</b>	<p>This enables the moderator to invite new people to join the moderation team. The invitation is sent to the user's reddit username via the internal messaging system.</p>
 <b>approved submitters</b>	<p>Enables the moderator to grant some special privileges to certain regular and trusted posters, such as allowing their posts to appear immediately, without going into the moderation queue for manual approval.</p>
 <b>traffic stats</b>	<p>This brings up traffic statistics for the subreddit, including the number of unique visitors per day/month/hour, the number of page views per day/month/hour, and the number of subscribers.</p>

 moderation queue	Some subreddits choose to hold new posts in a moderation queue until they are manually approved by a moderator. The moderation queue will also hold posts from domains or users that have been blacklisted in the Automoderator settings.
 reports	Posts and comments reported by users are shown here. The moderator can choose to ignore the report or act on it, e.g. by removing the material.
 spam	Posts which are suspected to be spam are shown here. Moderators can choose to approve posts which may have been wrongly identified as spam.
 edited	This enables the moderators to see any posts and comments that have been edited by another moderator.
 ban users	Moderators can use this function to ban users, permanently or temporarily. It sends a message telling the user they've been banned and can include an explanation of why.
 mute users	Temporarily prevents users from sending abusive personal messages to the moderators via modmail function.
 edit flair	Enables the moderator to set flair that can be seen on the site, decide whether users are able to flair their own posts or if only moderators can assign flair, and design available flairs.
 automoderator config or  get started with automoderator	Enables certain moderation functions to be automated. Automoderator configuration enables the moderator to alter settings. If it has not been set, 'get started with automoderator' provides information on how to begin.
 moderation log	This lists all the moderator actions that have been carried out on the subreddit by automoderator and the human moderators.
 unmoderated posts	This lists all the submissions that are waiting to be checked by a moderator.

These tools help the moderators with the day-to-day running of the subreddits, which can require several hours work a day, or only a few hours a month, depending on the amount of traffic a subreddit receives, how controversial the topics it covers are likely to be, and how strongly the moderators want to control what information is posted there. Some subreddits are heavily moderated, with tight rules about what can and cannot be posted; for example, r/science demands that all posts must include links to peer-reviewed science journals, and has a team of 1,500 moderators who enforce this; r/worldnews does not make this requirement and despite similar subscriber numbers (16 million in April 2017) has less than 100 moderators. Others require very little moderation as traffic to the site is sparse and/or uncontroversial.

The moderators I interviewed spoke extensively about the moderation tools available on reddit – which is what prompted me to explore them in more detail – and in general saw them as beneficial to the moderation process. The moderation log, which lists all moderator actions that have been carried out, is a particularly valuable tool for moderation teams who may be geographically separated from one another and in different time zones, as it enables them to see what has been acted on and what is still awaiting moderator intervention.

### ***Automoderator***

Perhaps the most useful tool in the moderator toolkit is ‘automoderator’, a system built into reddit that allows moderators to define a series of checks and actions that will be automatically applied to posts in their subreddit and which particularly helps with the policing role. Automoderator can be configured to automatically remove posts that receive a set number of reports from other users, comments or submissions containing certain words or phrases that have been added to its settings, such as known racist or sexist slurs or dubious alternative therapies, and to automatically accept link posts from trusted

domains and approved posters. This reduces the time burden on the moderator: rather than having to read every post to check its content, it leaves more time to focus on other aspects of site and community management, such as posting good content and replying to users' questions. Originally designed by a reddit user, u/deimorz, for use only on r/gaming, it was eventually taken up across the entire site (u/deimorz went on to work for reddit professionally for a period from 2013-2016, and remains involved with the site as a user). An excellent example of User-Centered Design, it is expected to improve as NLP (natural language programming) technology improves. An example of how Automoderator can be configured is shown below:

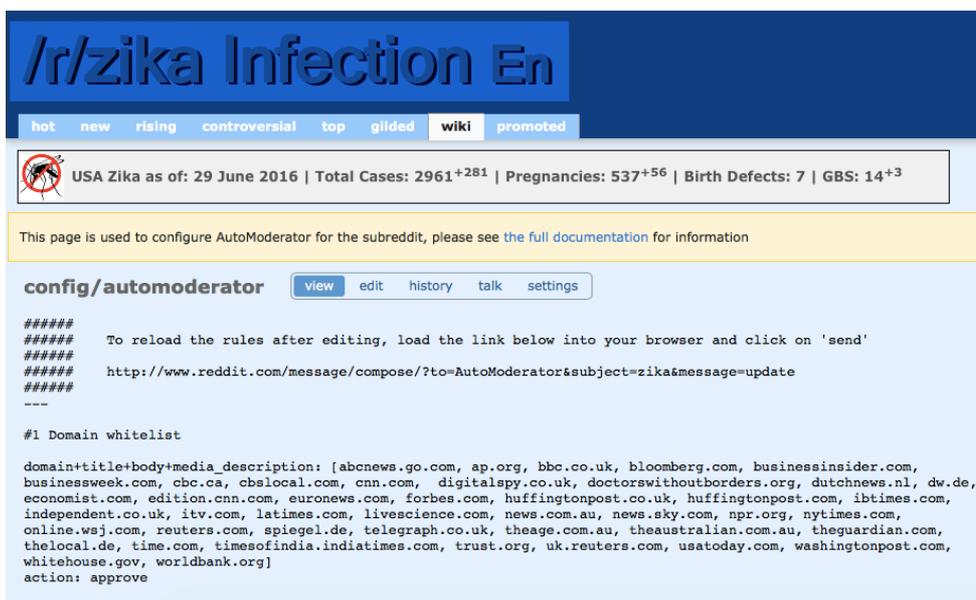


FIG 45: An excerpt from the automoderator configuration for r/zika, showing a list of domains from which links will automatically be approved.

Automoderator filters can be applied to all posts, 'quarantining' them in a moderation queue until they are approved by a human moderator, or just to posts containing certain suspect words, which can be predefined within the system, such as racist, sexist or religious slurs.

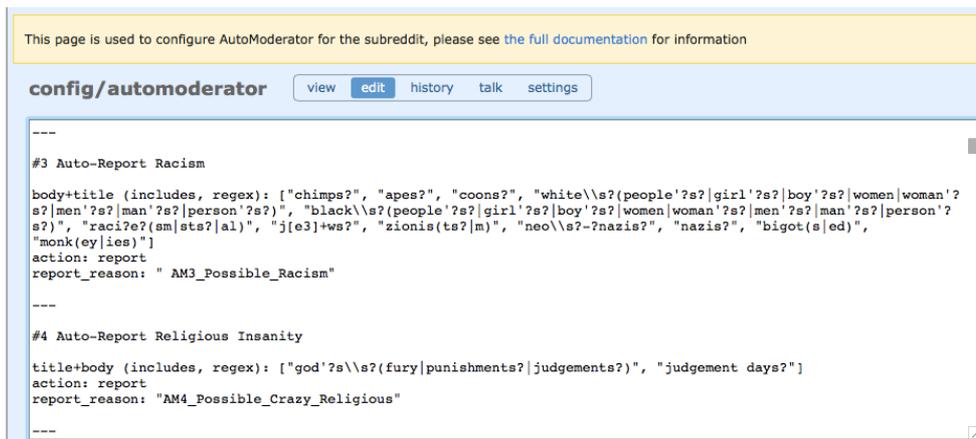


FIG 46: Another excerpt from the automoderator configuration for r/zika, showing a list of words that may indicate a racial slur or extreme religious views. Submissions containing these words will be held in the moderation queue until they are approved by a moderator.

Automoderator can be set up so that the forums will not accept submissions from new accounts (usually less than a day old), as experience has shown that these are often ‘throwaway’ – one-time use – accounts, which are more likely to be associated with negative and abusive behaviour (Leavitt, 2015), or newcomers who may not understand the site’s rules and positioning and are less likely than experienced posters to be making high-quality submissions. This does not prevent new users from viewing information on the subreddit, and therefore finding information that may answer health-related questions and concerns they have – only from contributing. The tools can also be used to ban – permanently or temporarily – users who violate the subreddit’s rules.

During a PHEIC, a subreddit could be configured so that posts from approved posters such as public health experts and officials appear immediately, while comments from less informed members of the public, who may spread misinformation, are held pending approval. This would enable subject-matter experts to help with the quality control and the socialising aspects of the forum while the moderators are tied up in management and administrative duties.

Automoderator is helpful for flagging problematic words that may indicate racism, sexism or poor quality information, for example any posts with ‘homeopathy’ in the title, or that mention therapies known to be particularly controversial. This is not necessarily because moderators always want to remove such posts, but in some cases just “so that they [can] be brought to our attention and we can join in with the conversation if necessary” [RM004].

This flexibility to ‘quarantine’ comments for review rather than remove them completely is also valuable as the inclusion of certain words does not automatically point to a problematic post: RM015, a moderator of r/books, reported that posts which were automatically filtered out for containing racist or sexist terms were often found to be discussing the use of these words in a literary context or in a poem. The moderation queue function enables valuable submissions to be filtered back into the subreddit while ensuring that poor quality ones can be discarded.

How much the moderators use automoderator, and how they configure it for their subreddit, can strongly influence the quality of the information that will be displayed, but the automoderator tool is only as good as the moderator who sets it. Some forums may prefer to moderate manually, or not feel that the moderation tasks are a drain on their resources: r/starwars did not use Automoderator until shortly before the release of a new Star Wars movie in December 2015, despite having more than 300,000 subscribers. The moderators only began to use it on the suggestion of a moderator from r/ebola, to deal with an anticipated (correctly) influx of racist comments relating to the casting of a black actor in a lead role, cutting and pasting r/ebola’s anti-racism configuration for use on the forum. The ability to copy automoderation configurations from other subreddits when they are needed, rather than having to programme and develop them from scratch, eases the need for the forum moderator(s) to have extensive technical skills themselves.

The tool's efficacy comes mainly in the timesaving it affords the moderators and its use seems generally tied to the size of the forum and the amount of traffic it receives, though this is not a straightforward relationship. It can also be influenced by the amount of time the moderators are prepared to spend on general housekeeping tasks: some small forums use it extensively, while other larger forums hardly use it at all. The important point is that it is there, but using it well does require some programming skill and/or experience, though other moderators can and will help with this if requested – and how this is achieved will be considered later in this chapter.

### ***Additional tools***

Moderators can also use additional tools to automate actions on their subreddits, for example to rotate graphics that appear at the top of the subreddit frontpage, to automatically remove posts from suspected spamming domains and to schedule AskMeAnything sessions – which could be particularly popular during a health emergency. More experienced technical editors also reported drawing on tools such as Notepad++ for text editing [RM008]. Effective use of these tools, which are part of the channel characteristics of the subreddit, helps to improve the surface credibility.

Moderators reported wanting some functions, not currently available, that could be considered for future systems, such as the ability to be able to search easily through previous posts and comments (which moderators have been told by the Admins is hard to implement and is unlikely to be introduced any time soon) [RM015]. Natural language processing (NLP) that is more adept at picking out bad information than the current system of blacklisting certain words [RM018], such as the commentIQ software used to sort comments made to *The New York Times* (Park et al, 2016), was also mentioned as a potential improvement to the current toolkit.

### ***Moderation tools and influence on content quality***

How the moderation tools are set up clearly has a strong influence on the quality of submissions that appear on the subreddit. Posters and domains can be banned outright, or the subreddit can be configured so that no posts appear on the public-facing areas of the forum until they have been approved (or not) by a moderator. Such quarantining can be applied at times when an influx of poor quality information is expected, but is most often applied in forums which are prone to persistent offensive or abusive posting, which is not unique to reddit. Golder and Donath (2004) reported similar action by a breakaway group from the Usenet community alt.computer.consultants, which wanted no part in hostile discussions about IT jobs being outsourced to India that had become prevalent on the site. They set up alt.computer.consultants.moderated, a separate forum on which a volunteer moderator had to approve all posts before they appeared. On reddit, r/zika holds all posts until they have been approved by a moderator, to guard against racism and conspiracy theories that proved to be an issue on r/ebola at the height of the outbreak.

### ***Disadvantages of technical requirements***

There are some disadvantages associated with the technical tools available to reddit moderators through the platform, however. Some reddit users, who would otherwise be likely candidates to take on moderation, can be put off by this technical side. One of the people I interviewed [RM006] was not formally a moderator – i.e. recognised by the system and offered some system privileges not available to non-moderators – but was recommended for interview by one of the actual moderators of the forum as they considered that he acted as a moderator and would be more able to answer my questions in a way that would be useful to my study. In effect, the formal moderator had strong technical skills, and performed the Manager and Engineer functions of

the moderation requirements, but delegated the subject matter expertise and socialising functions to regular posters.

The poster I interviewed [RM006] was a regular and experienced forum user who had been invited to join the moderation team but declined the offer as he did not feel he would be able to deal with the technical requirements. As he was not being expected to perform any of the Engineering or Administration functions that require the privileged access to additional systems only available to formal moderators, there was no need to push him on his preference to decline. Without it, he can still post useful content, provide informed answers to questions asked by other posters, help newcomers to feel welcome and report concerning content to moderators (which he often picks up quicker than they do, as he is more active on the forum, for longer periods of time, than them). While he cannot remove bad information himself, he can quickly alert a moderator that action may be needed. This further supports the point made earlier about quality control being the role of all reddit users; the moderator has additional roles on top of this.

### ***Learning technical skills***

Opinions on how easy the technical aspects were to pick up differed. RM008 felt that, “Reddit has a fine platform which is dead simple to work with” and RM011 reported that, “[The technical skills] are relatively easy to pick up. I have an IT background, I have programming experience, so I find it relatively easy. Several years ago, it was more difficult and actually the Administrators of reddit and the other reddit users, have continued to make all these improvements, to make it ever easier to moderate”. Others felt that the tools “can be a little bit intimidating for people to figure out. There are places you can copy and paste ... but it can take a few days to properly figure out”, reported RM015, who “would like to see it easier for moderators to be able to

put restrictions for posting – not to block all posts – without using automoderator”, though she also acknowledged that, “If you know you need to find someone [with technical skills] you probably can. To find someone who can help you design your CSS or provide your graphics and things, and to find someone who can use automoderator and help you with that is usually not super hard.” There are even subreddits dedicated to helping less technically-adept moderators, such as r/modhelp, where moderators can either ask how to configure a setting or appeal for someone who can do it for them.

It is important to state, however, that the existence of these tools is only visible to the moderator once they have taken on that role. A prospective new moderator may be entirely unaware of their existence, or the need for them, beforehand. For this reason, it is valuable to spend some time explaining how one can become a moderator on reddit, and at what stage of the process the requirement for technical skills starts to become apparent and important.

## **7.5 Becoming a moderator**

Any registered reddit user can become a moderator, and there are three routes through which this might happen:

- Be invited to join the moderation team of an existing subreddit
- Offer yourself as a moderator to an existing subreddit
- Set up a new subreddit; by default, the person who sets it up becomes Lead Moderator

Any registered reddit user is free to set up a new subreddit on any topic they choose, including topics for which subreddits already exist, so long as the name of the subreddit is not already in use (for example, the existence of r/SARS would not prevent a subreddit called r/SARS\_outbreak being set up).

Regular users on one forum may decide to set up a more focussed aspect of the topic they were discussing, or a subreddit discussing the topic from a different angle: r/chronicpain was set up by regular posters on r/opiates who wanted to discuss pain-relieving drugs for medical conditions but found discussions on r/opiates tended towards recreational drug use by “junkies” [RM005]; others may want to set up a subreddit on a topic not currently being discussed on reddit at all, which was how r/ebola was originally started, by a student who had learned about the disease in class and was interested in creating a space where he could discuss it further. Once the user has set up the subreddit, they become the forum’s Lead Moderator. The user must have some experience of reddit to do this: their account must be at least 30 days old, and they must have accumulated a determined amount of positive karma (the exact amount is not specified and is officially known only to the reddit Admins but is generally accepted to be 100). This requirement may disadvantage a subject-matter expert seeking to set up a new subreddit in the very early days of an outbreak, but can be over-ridden by the Admins.

The process of setting up a new subreddit is reasonably simple and does not require much in the way of technical skills. The prospective moderator clicks the ‘create your own subreddit’ button on the reddit front page and a simple content management system then enables them to set up the subreddit, give it a name and title; a description; fill in additional information that will appear in a sidebar on its homepage; specify which language it is (expected to be) in; and to set access – public, restricted, private (explained in Chapter 4, p131) or gold only. There are also options for setting up what type of content can be submitted (for example, the moderator may choose only link-, or only self-posts), the strength of the spam filter, and some basic background colour design elements. The content management system is simple to use and requires no more technical expertise than is required for using reddit in general. In other words, any reddit user should be able to do it.

Once a subreddit has been set up, it cannot be deleted, nor the name changed. If the moderator who set it up no longer wants to moderate it, they can remove themselves as a moderator and – if there are other moderators – one of them may take it over or, if there is only one moderator or no-one wants to take it over, it will become ‘abandoned’ and available for any other reddit user to ‘adopt’ by messaging the Admins and asking to do so. Moderators who are intending to abandon their forums are encouraged to post their intentions on subreddits such as r/needamod, so that other users know and may decide to take them over. Once users have set up a subreddit, they generally will moderate it themselves, at least in the initial period after setting it up, and are free to add others to the moderation team at any time. Moderators who set up a subreddit themselves may be unaware that any skills other than subject matter interest are needed. Their level of technical skills may be no higher than that needed to use the content management system to set up the subreddit.

Secondly, any registered reddit user can be invited to join an existing moderation team. This is most likely to happen to active posters who the existing moderators consider to be submitting good quality information, either in the subreddit needing a moderator, or elsewhere on reddit. The invitation can be made directly by the existing moderator(s) to the poster, or the moderator(s) may indicate they are looking for help through a general post on the forum, to which the user replies and is accepted. If existing moderator(s) need help with a specific aspect of the subreddit, or advice on a specific issue, users may be invited to join a subreddit they have not participated in before based on specific subject matter expertise, technical skills or context skills they have. For example, r/zika sought out Portugese-speaking moderators, as many early cases affected Brazil, a Portugese-speaking country; r/ebola moderators reached out to people with experience of moderating larger forums when traffic significantly increased at the height of the outbreak.

TABLE 28: *How those interviewed for this study became moderators*

INTERVIEWEE	MODERATED BEFORE THIS FORUM?	INVITED REPLIED OR APPROACHED?	LEAD/ADDITIONAL OR TOOK OVER?	HOW RECRUITED
001	N	Invited	Joined existing subreddit team	Active user of subreddit
002	N – but now does others	Invited	Joined existing Subreddit team	Active user of subreddit
003	N	Part of professional role	Lead	Set up forum
004	Y	Replied to call	Joined existing subreddit team	Active user of subreddit
005	Y	Set up one; then replied to call on another	Set up; joined existing; took over abandoned	Active user of subreddit
006	N (not moderator)	Invited but declined	n/a	Active user of subreddit
007	N	Invited	Joined existing subreddit	Active user of subreddit
008	N	Joined semi-abandoned; invited	Joined existing subreddit	Active user of subreddit
009	N	Set up	Lead	Active user, other subs
010	Y	Invited	Joined existing subreddit	Active user of subreddit
011	Y	Set up, and took over abandoned	Lead	Active user, other subs
012	Y	Invited	Joined existing subreddit	Active user, other subs
013	N	Took over abandoned	Took over abandoned	Active user, other subs
014	N	Invited	Joined existing subreddit	Active user, other subs
015	Y	Offered help	Joined existing subreddit	Specialist
016	Didn't say	Didn't say	Didn't say	Didn't say
017	Y	Responded to call	Joined existing subreddit	Specialist
018	Y	Invited	Joined existing subreddit	Active user of subreddit

Being invited to join an existing moderation team had been the route into moderation for 13 of the moderators I interviewed, and five [RM006, RM008, RM013, RM015 and RM017] reported that they had also recruited others based on specific skills or experience needed. These moderators are likely to have subject matter skills, and some context skills, but may or may not initially have technical skills.

Thirdly, reddit users who think they have knowledge and/or skills from which an existing subreddit will benefit can offer themselves to the forum as a moderator, usually by sending a personal message to the existing moderator(s) or by responding to a call posted on the forum. These users are likely to have strong context skills: they have demonstrated experience of the forum by being active on it, have an awareness that the moderator needs help they may be able to provide, and have the subject knowledge, technical expertise or experience they are offering. Three of the moderators interviewed [RM004, RM005 and RM017] had become moderators of a specific forum in this way; two had moderated other forums before, where they had demonstrated the skills required.

## **7.6 Moderation teams**

The above highlights an important characteristic of reddit forums: they tend to be moderated by teams, rather than by a single moderator. The obvious requirement for moderation teams is to share the workload on a busy forum and, in the case of the subreddits whose moderators I interviewed, the more popular a subreddit was, and the more subscribers and traffic it attracted, the larger its moderation team tended to be, shown in FIG 47. The ability of the reddit system to support a team of moderators is a definite advantage of the platform, as it enables the forum to scale up and for the workload to be distributed – which is particularly useful when the moderators are volunteers.

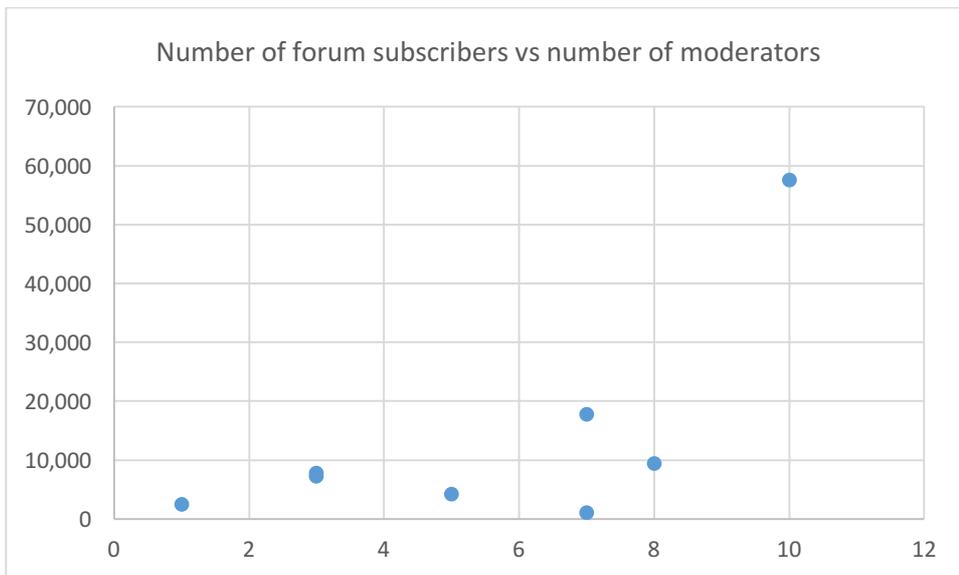


FIG 47: *There is a general trend of an increase in the number of moderators as the number of subscribers increases, for the subreddits from which the moderators I interviewed came. A marked increase in the size of the moderation teams comes at around the 10,000 subscriber mark.*

Workload is not always equally distributed between the moderators, however; some clearly do more than others. For example, of 1439 moderation actions recorded on the moderator log of r/zika between its creation 6<sup>th</sup> January 2016 and 1<sup>st</sup> April 2016, 87% were carried out by a single moderator.

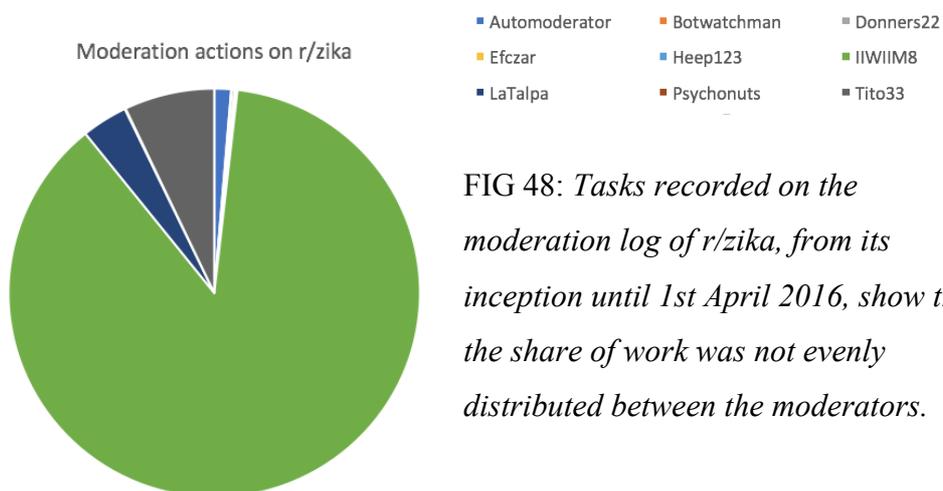


FIG 48: *Tasks recorded on the moderation log of r/zika, from its inception until 1st April 2016, show that the share of work was not evenly distributed between the moderators.*

From the interviews I conducted, it appears that when moderation teams expand, they do not tend to just distribute the workload equally. More common is that moderators within the team take on different roles, some of which are very specialised – such as the programming skills to set automatic functions or the graphic design of its front page. RM017 described himself as “the graphics guy ... just making nice pictures to pretty up the subreddit”, and RM015 also described one of the subreddits she moderated as having a dedicated graphics person, whose sole job was to ensure that the banners on the front page rotated regularly. “One of the guys does all the CSS for the subreddit,” reported RM005. This was not only true of technical skills: it also included moderators whose role was to deal with certain types of troublesome behaviour likely to be encountered, such as configuring automoderator and the spam filters to guard against hate speech. RM015 reported sending the moderators of a forum which had been experiencing racism, “a bunch of rules [to configure automoderator] and they didn’t even know how to add any of the rules, so they just added me as a moderator and said, ‘do whatever you want’ .... Now, I usually just send rules [to other subreddits] and if they don’t know how to put it on, they’ll modmail me for an hour and I’ll add them and then go. If they need more help with that later, they can just let me know”. Smaller forums, which may not need a large moderation team but whose moderator(s) may occasionally need some additional skills (such as graphic design) temporarily, can ‘borrow’ a moderator for this when needed.

An interesting point in this examination of the differentiation of roles is the role of the Lead Moderator. For the subreddits whose moderators I interviewed, the Lead Moderator was not necessarily the most active, usually because their interest in the forum had waned over time, but they may remain on the moderation list and hold additional privileges to the other moderators. This was not necessarily considered a problem, as it provided an independent arbitrator who could come in and solve disputes when needed. The process of

group moderation appears to work as a collaboration between peers, with the Lead Moderator rarely invoking their privileges. Only if they are completely inactive can it cause problems for the remaining moderators, but in such a case they can appeal to the Admins to have the status transferred to them. From the reports of the moderators I interviewed, this is generally granted.

The moderators interviewed in this study spoke about the team aspects of moderation favourably; there was little experience (or even hearsay) of serious disagreements between moderators. It was more usual that moderators would work together to collectively overcome a problem that may have been more difficult for a single moderator working alone, such as dealing with a rude or abusive user, particularly following a ban. Moderators could back up one another's decisions, thus providing moral support and further justification of action taken, as it could be presented as a joint rather than individual decision.

The 'message the moderators' tool is particularly useful here, as it enables the moderation team to have a discussion around the message, which they will see in the modmail forum, and can discuss between them how best to respond. It is also possible to configure modmail so that a reply made to a user comes from the moderation team, rather than from an individual moderator, which can deflect individual abuse that may result from a ban or removed comment that is perceived as unfair.

Moderation teams communicate with each other in a variety of ways, using the modmail system within reddit itself and often setting up moderator-only private or restricted forums where the moderators of specific subreddits can discuss forum administration and issues. These include [r/ebola\\_modmail](https://www.reddit.com/r/ebola_modmail) ([https://www.reddit.com/r/ebola\\_modmail](https://www.reddit.com/r/ebola_modmail)) and [r/armyofscience](https://www.reddit.com/r/armyofscience/) (<https://www.reddit.com/r/armyofscience/>). Moderators also reported using additional communication platforms to those available through reddit, particularly ones that enabled live chat for when issues needed to be discussed

and resolved quickly, such as Slack (<https://slack.com>) a messaging app; Discord (<https://discordapp.com>) a similar system particularly favoured by gamers; Google Hangouts and IRC (Internet Relay Chat) channels. It was common for groups to have one realtime communication method (e.g. Google Hangouts) alongside one message board system to allow for time zone differences of moderation teams that are geographically dispersed. Building such a system into a future platform may prove valuable, particularly during health emergencies affecting more than one country or timezone.

Few moderators knew, or expressed much interest in knowing, one another offline – the system seemed to work perfectly well online without the need for face-to-face interaction. “A certain degree of ‘message in a bottle’ is present, where answers and ideas come sporadically. The rule ... is ‘real life takes priority’; what might be viewed as addressing a pressing need on one side may be an untimely interruption on the other”, said RM008. The ability for teams of moderators to communicate without having to reveal anything about their offline identity was considered a definite advantage. Moderators did not necessarily even know the gender of others on their moderation team.

## **7.7 Moderation requirements and Context/Community Shift**

There seems to be a different emphasis on the need for, and consequently the time required to focus on, different moderator roles at different points in the subreddit’s life and development, consistent with the concepts of Context Shift and Community Shift discussed in Chapter 5. When the subreddit is first created, “It has a plain look, few if any embellishments. It lacks the qualities of organisation found in developed sites. This is when moderators make major decisions: how it’s going to look, what the user guidelines are, what is and isn’t allowed; controlling appearance, who’s going to do what tasks. Automoderation, flair, rules, style sheet and subreddit settings are addressed

and fleshed out” explained RM008. Once the forum is up and running, this turns into daily maintenance and upkeep, approving or removing posts and comments that appear in the moderation queue.

Understanding what is needed when, and how best to provide it, leads me into discussion of the third skill set needed by the moderator: Context Skills.

### ***Context skills***

A new and inexperienced moderator may not be aware of the technical tools available to help with the role or that they may need some technical skills to make the most of them. Nor may they be aware of problems such as spam, racist groups, conspiracy theorists, purveyors of dubious alternate therapies, or other challenges the forum moderator may face. More experienced moderators – particularly those who have worked across more than one forum – are not only aware of these challenges, they also know of other community members with skills that can be called on when needed to deal with them.

The tendency for subreddits to be moderated by teams, and for many new moderators to join existing, already experienced teams, enables the new moderators to learn the skills that are less immediately obvious ‘on the job’ with support from their co-moderators. Not only is reddit a vast repository from which knowledge can be drawn, it is also a vast repository of reddit users who, between them, already have all the technical and context skills a new moderator needs to moderate their subreddit efficiently. This knowledge is held by the community and can be retrieved from it when needed just as subject matter expertise is. This provides the moderator with context agency to configure automoderator and other tools; to find subject matter experts when needed; locate those with experience of dealing with problematic posters and problematic posting behaviour; and with experience of how to deal with

challenging issues such as periods of rapid growth, including what is likely to trigger them so that they can be anticipated and planned for. A reddit user and moderator will build up context skills from personal experience, but a significant benefit of reddit is access to the wider community of reddit and the social capital it provides. In this way, the moderation teams contribute to the ability of the reddit user to harness the Collective Intelligence of the platform, particularly through moderator metacommunities.

### **7.8 Moderator metacommunities – reachback into reddit crowd**

The reddit *Crowd* contains a metacommunity of experienced moderators, who can be sourced when needed and who can come forward to offer help to less experienced moderators who may not even be aware they need it. This metacommunity cuts across the subreddits and has emerged out of the structure of reddit itself. Reddit maintains subreddits dedicated to moderation issues – including r/modhelp, which has over 8,500 subscribers (8,850 on Dec 31 2016), r/findamod, r/modclub, r/csshelp and r/modnews, where announcements about new moderation tools and changes to site structure and policy are announced. On 31<sup>st</sup> December 2016, r/modhelp listed 18 other subreddits dedicated to moderation issues in its sidebar information. These provide access to the collective experience of all the reddit moderators, irrespective of subreddit topic focus, and can provide advice and support to less experienced, less tech-savvy or less subject-matter expert moderators.

This moderator metacommunity is particularly interesting in light of the finding by Buntain and Golbeck (2014) that reddit users rarely exhibit significant participation in more than one community. Moderators in contrast seem to participate across a variety of subreddits, not all of which are thematically linked, particularly in the case of those moderators who take across specific skills such as CSS, design, automoderator configuration and

rule-setting. Further research into these ‘specialist moderators’ would be particularly valuable, but was outside of the scope of this study. Also, when moderators have been studied, it has largely been in isolation from one another with little consideration of how a team of moderators might work together or how well, together, they may meet the criteria identified by Woolley et al (2015), of moderate diversity in cognitive styles and approaches to problem solving that is best suited to enabling a Collectively Intelligent group. This, too, would be a valuable area of further study.

## **7.9 Conclusions**

The process of moderation on the reddit forums examined for this study is multi-faceted and more complex than it first appears, to both users of the site and to prospective new moderators. Effective moderation includes both human systems and automated processes, for which moderators need to have (or swiftly acquire) not only subject matter expertise, including scientific and medical knowledge and experience of the health condition, but also technical skills including CSS programming and graphic design, plus context skills related to online behaviour, internal and external challenges commonly experienced by the reddit platform, and awareness/understanding of the wider reddit community.

Reddit’s *Collection* of subreddits and of moderators who constitute a metacommunity that congregates within communities such as r/modhelp – affords easy location of potential moderators with a variety of skills. In this way, the skills required for moderation are held collectively and can be passed between moderators and subreddits as needed – a Knowledge Age version of the cooperation discussed by Tomasello (2009a) and other anthropologists as being a key evolutionary requirement of human cognitive ability. An experienced reddit user, seeking to set up a new subreddit quickly, can configure a new forum and assemble a suitable moderation team. In addition,

when experienced moderators notice new or rapidly expanding subreddits, they are able to anticipate the challenges it might face and offer their help accordingly. All this, is however, is dependent on the necessary contextual experience to understand the processes and how they work.

A new moderator, who is most likely to emerge from an existing subreddit where they have been posting regularly may be entirely unaware of this but the structure of reddit overcomes the challenge in two ways: the tendency towards moderation teams rather than a single moderator helps to ease newcomers into the position gradually, learning the skills they need on the job, and the moderator metacommunity provides a deeper reachback into the reddit *Collection*, where communal knowledge of the moderator process resides.

This warrants reflection on the point made in Chapter 5, that while chronic condition subreddits are *Support* subreddits, PHEIC outbreak subreddits are *Information* subreddits, and their characteristics are different – which I will explain further in Chapter 8. A new subreddit, set up during the early stages of an outbreak, may not allow sufficient time for moderators to build up the experience they need, nor for them to learn how to make full use of the automation features, before the number of cases leads to a corresponding increase in the number of subscribers, traffic and posts. This may become a challenge, particularly if the relative newness of the subreddit means that it also lacks an experienced user base from which to recruit additional help. Chapter 6 showed that posts on an outbreak subreddit may be more likely to contain poor quality information more often than on chronic condition subreddits, making quality control across the outbreak subreddits harder for the moderators to maintain. In the Far-at-Risk and Near-at-Risk stages, ‘hard’ scientific and medical knowledge is needed to answer questions and deflect misinformation from poorly-informed posters: the quality control role may be most needed at a time when the managerial, administrative and technical burdens on moderators are also at their highest.

Help is available, however. In the early days of r/ebola, the relatively inexperienced moderators found that help was offered freely by, “redditors with deep experience and valuable knowledge of the workings of reddit. Soon thereafter, all the problems [...] were resolved” [RM008]. Exactly how this happened, and what lessons can be learned from it, are explained in the next chapter, in which I will examine how r/ebola was utilised during the early stages of the crisis and scaled up as the outbreak spread. If the time required to build up experience is seriously curtailed, it is interesting to know if – and how – this might be compensated for by other characteristics of the platform, either those that exist already on reddit, or those that might be incorporated into future PHEIC platforms. I will now explore these themes in Chapter 8.



## 8: HOW SUBREDDITS SCALE UP IN A PHEIC

---

### 8.1 Introduction

A Public Health Emergency of International Concern (PHEIC) may involve a newly emerged disease, or a disease which has previously been of interest only within a very specialised scientific or medical community. Such diseases may not have a subreddit dedicated to discussing them when an outbreak begins, or any subreddit that does exist may have just a handful of subscribers. On 25<sup>th</sup> March 2014, the day the 2014-16 Ebola outbreak was officially recognised by WHO, r/ebola had just four subscribers and had received only one post in the year since it had first been established, made on 19<sup>th</sup> May 2013. Neither r/zikavirus (established 6<sup>th</sup> January 2016) nor r/zika (established 27<sup>th</sup> January 2016) existed at the time the outbreak was declared a national public health emergency by Brazil, on 11<sup>th</sup> November 2015, and were set up only shortly before it was officially declared a PHEIC on 6<sup>th</sup> February 2016. The r/MERS subreddit dedicated to Middle East Respiratory Syndrome (MERS), often considered to be a likely candidate for a future PHEIC, was created on 28 April 2014 had 46 subscribers on 1<sup>st</sup> January 2017.

This may pose some challenges to reddit's utility in a PHEIC, if I was correct in Chapter 7 in pointing towards moderator experience – of the forum they moderate and of reddit in general – as a key factor in quality assurance. While a subreddit on any topic can be created from scratch and be ready for use within minutes, the interviews I undertook with the reddit moderators, the record of conversations in the r/ebola and r/zika modmails, and the moderation logs to which I was given access, suggest that building and maintaining a *good* subreddit is somewhat more complicated. In this chapter, I will explore how new subreddits are set-up, how they grow as they attract more traffic, and how subreddits for existing health conditions and those for newly emerging ones

compare. I examine how the moderator process works in this context and the influence on quality it has, or may have, at various stages of that process.

In theory, a subreddit dedicated to a PHEIC starts in the same way as any other subreddit. A reddit user chooses to start it, and becomes by default its lead moderator. They may or may not add additional members to the moderation team, and may or may not choose to promote their subreddit to other users, which may depend on whether they want to keep the community small or deliberately encourage it to grow. None of the moderators I interviewed were deliberately trying to limit the size of their community, and the only two [RM004 and RM005] who expressed any sentiment against wanting the forum to grow were both chronic condition moderators who regretted that new subscribers generally reflected new diagnoses: “Generally you’re there if you’re dealing with [the condition] yourself, or you have a family member or friend who you’re trying to figure out how to help, so I don’t like seeing that subscriber count go up. If it goes down, I’m happy because that means some people are getting better” [RM005].

Once a subreddit is established, users can find it by searching on keywords relating to the topic it covers. There are also ways to promote a new (or existing) subreddit. Those described by the moderators I interviewed including actively searching for related terms and, if any posts using that term had been used in another forum, sending a personal message to the poster, or replying to their post with a comment, informing them of the new subreddit, directing them towards it and encouraging them to contribute. Another method was to apply to be ‘subreddit of the day’ which, if accepted, provides a banner advertisement on the Front Page. Health subreddits tend to start out reasonably small and grow a subscriber base gradually over time. The graphs in FIGS 49-51, from reddit metrics ([www.redditmetrics.com](http://www.redditmetrics.com)), show the growth curves for three health condition subreddits (r/cystic fibrosis, r/diabetes and r/asthma).

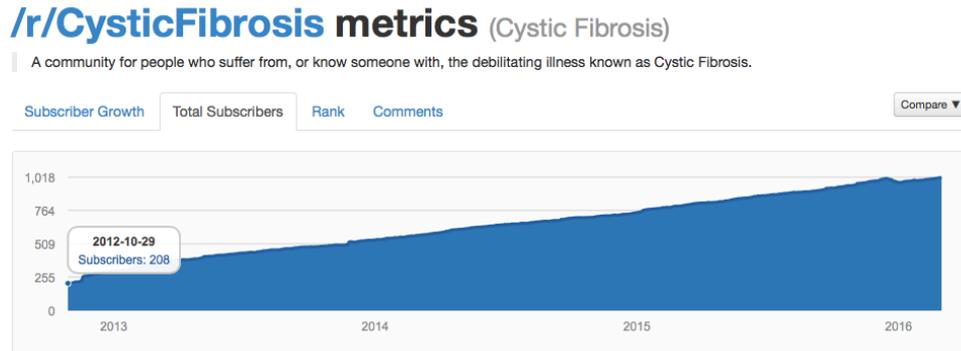


FIG 49: Growth of subscribers on r/cystic fibrosis. Note the apparent dip in subscribers at the beginning of 2016 (and also visible in FIGs 50 and 51): this is due to reddit-wide site administration that removed deleted accounts from subscriber lists. Previously, reddit users had remained counted on subscriber lists even if their account was deleted.

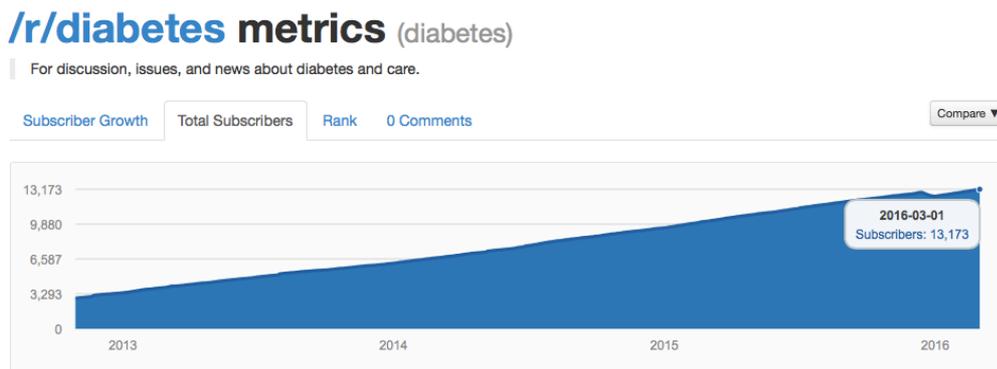


FIG 50: Growth of subscribers on r/diabetes.

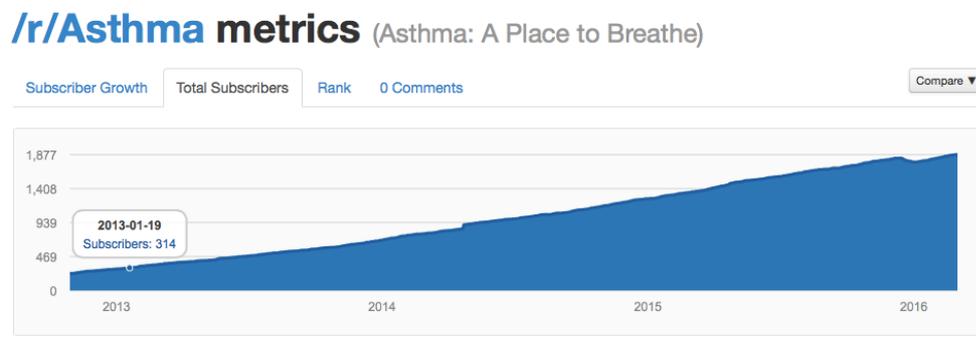


FIG 51: Growth of subscribers on r/asthma.

The three subreddits shown here follow a consistent pattern of slow but steady growth over time, a pattern consistent with subreddits in general – r/news, r/science and r/books all show similar growth patterns for example. It is also common to health forums across other platforms I looked at, including HealthUnlocked and Patient. As the HealthUnlocked moderator [RM004] described: “it grew fairly steadily for the first couple of years and now it’s growing even more quickly. We’re nearly at 10,000 members” [after five years; r/ebola gained this increase in two weeks in October 2014].

Online communities develop and settle down gradually, absorbing additional new members as they arrive with relative ease. The steady trickle of new arrivals adapt to the existing norms and conditions of the forum, and are integrated into it, helped by the socializing behaviour of the moderator and existing members (Hsieh et al, 2013) and, on reddit, the guidelines for behaviour that are often displayed on the subreddit’s first page or as a ‘stickied’ post, which remains permanently in place at the top of the forum.

The way a subreddit develops is not always exactly as intended, however: the original moderator of r/publichealth used the subreddit as a platform for conspiracy theories while a subsequent lead moderator [RM013] had a “grand idea of it being a place of discussion [on factually accurate public health topics]”. It has evolved instead into a forum where public health professionals discuss career advice and where aspiring public health students discuss the merits of various courses and universities. This was the aim of neither its creator nor later ‘owner’, but has been shaped by its community of users. It is interesting to note that while its users felt compelled to steer it away from becoming what they perceived as a negative positioning – a platform for conspiracy theories – those same users offered little resistance when others steered it from the positive positioning they had envisaged towards one that was different, but no less valuable than their plans for it.

Gradual growth of a subreddit is reasonably easy for the moderator to manage and if the community does grow to a size where help is needed, the forum itself provides a perfect recruiting ground. Twelve of the moderators interviewed [RM001, RM002, RM004, RM005, RM006, RM007, RM008, RM010, RM014, RM15, RM017, RM018] had either been recruited, or had recruited others, from subreddits in which they were already active as posters.

This has the advantage that the moderator is familiar with such users and can gauge how well they are likely to fit in with the existing moderation team:

“[W]hen you’re moderating a community, eventually you start to notice the regulars, the guys who are always trolling or, you know, just messing around, and then the nice guys, who explain in detail a lot of things. The moderators, when we agreed we needed new moderators, first off, recommendations – do you know anyone in the community who you think would be a good moderator?” [RM017].

A second consideration, however, is the technical aspects (which have been described in Chapter 7). After its initial set up, a subreddit can be developed in different ways. Moderators may decide to change the look of the forum from the basic default design to a more elaborate one (see FIGs 52 and 53); add rules that set out what behaviour is and is not expected from users; configure Automoderator to filter out certain posters, domains or terms; and make other changes to the default settings provided by reddit.

Doing this takes time, and particularly for less experienced moderators, can create challenges if the subreddit needs to be set up quickly. While some of the modifications – such as graphics and design – are largely cosmetic and are like to have little if any influence on surface credibility (as the default setting already looks professional and is easy to navigate) others, such as to Automoderator, are more important.

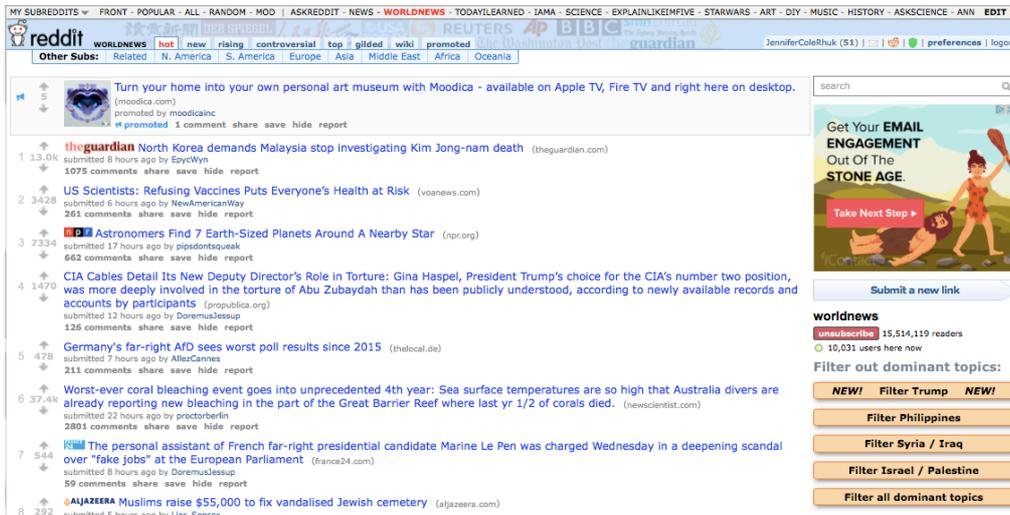


FIG 52: *r/worldnews* has not altered its design from the default setting, despite being a large subreddit with more than 15 million subscribers in Feb 2017.



FIG 53: *r/books* design has been modified to be more graphic. Some subreddits have moderators who work specifically on design and graphics.

How, or if, moderators configure Automoderator to filter out posts and comments containing certain words, how they develop and assign flair, set rules, modify the page's wiki, and make any other 'housekeeping changes' are far more likely to impact on the quality – and perceptions of quality – of the subreddit by influencing surface and message credibility. These require some technical expertise and experience.

When new subreddits are set up, a gradual increase in new subscribers affords the moderators time to configure their settings, learn what modifications might be needed and how to make them, and decide on the tone and direction the subreddit should take. They may decide to encourage self-posts over link posts, or vice-versa. If the early days of the sub are relatively quiet, moderators have time to make major decisions about its future; how and if to modify its appearance; what user guidelines to set; what is and isn't allowed. They can decide amongst themselves if a differentiation of tasks is needed and, if so, who is going to do what. Automoderation, flair, rules, style sheets and subreddit settings are addressed and fleshed out during this period.

## 8.2 The PHEIC subreddits

The PHEIC subreddits display a different growth pattern, however – rather than a gradual, steady increase in the number of subscribers, they experienced a rapid increase at the beginning of the outbreak, which then levels off and eventually declines when the outbreak is over, illustrated by the subscriber statistics for r/ebola and r/zika in FIGS 54 and 55. It is therefore worthwhile to explore the differences between the ‘lifecycle’ of a PHEIC subreddit and those of other health-related subreddits.

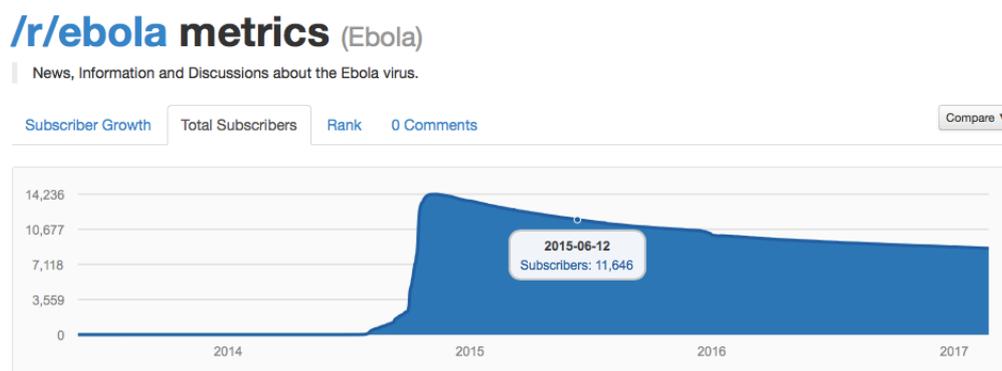


FIG 54: *Subscriber growth on r/ebola, which increased a thousandfold over in two weeks in October 2014. The increase in subscribers coincides with the rapid increase of cases in West Africa during the same period.*

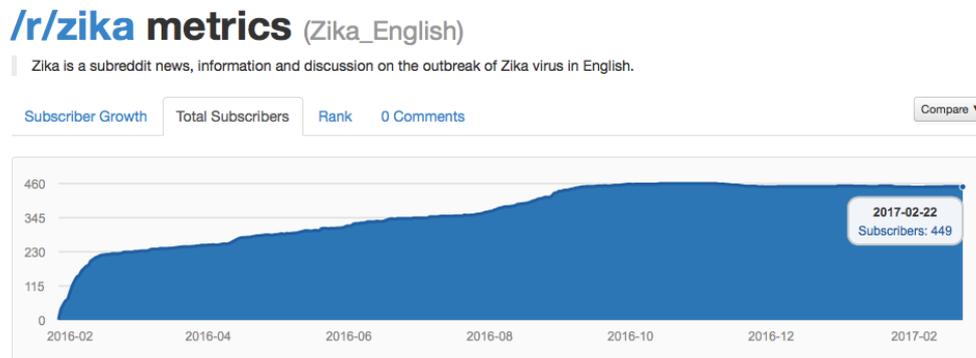


FIG 55: *Subscriber growth on r/zika, which grew rapidly over two weeks in February 2016 as news stories of the outbreak appeared in international media; around 50% of the subreddit's total subscribers, as of Feb 2017, joined in that short period. Zika was declared a PHEIC on 6 Feb 2016.*

One of the main challenges that might be faced during a PHEIC is whether to start a new subreddit, or 'piggyback' on an existing one. r/ebola, the largest of the PHEIC-dedicated subreddits, existed before the 2014-15 West Africa outbreak, and provided an obvious home for the discussions; the r/zika and r/zika\_virus subreddits were both set up after the zika virus outbreak started.

Setting up a new subreddit has the advantages of enabling the initial moderator(s) to set clear rules and intentions for the site from the beginning, and while these may change organically as the outbreak continues, they provide an initial direction for moderators and users to follow. For example, a new forum may decide to flair posts as originating from media sources (which may be more speculative and sensationalist) with the flair 'Media', and those from peer-reviewed academic journals as 'Science/Medicine' or identify those from international authorities such as CDC and WHO, enabling users of the forum to make quick credibility assessments on the likely quality of their content. A pre-existing forum may have already made decisions on whether it will accept posts from media sources, or only from peer-reviewed journals,

and be loath to alter its position. Unless there is an already established home for discussions, this setting up period can be, “a conflicted time [... as] moderators with no common history come together to create an entity” [RM008]. A moderation team is needed, but there is no time to enable one to form naturally in the way that seems to work well for existing subreddits. There are no experienced users to elevate to moderator status, and even the moderators may not have experience of moderating a forum on that specific topic – or in fact of moderating a forum at all.

PHEICs do not have the luxury of time afforded to other subreddits, and it is therefore worth understanding how PHEIC subreddits emerge, how they are set up and how the moderator teams that are so crucial to their quality control are recruited. It is also worth looking at the specific events that trigger a sudden upsurge in interest in a subreddit and drive a large increase in traffic and subscribers, as this may provide some pointers as to how the process may be managed and if (and how) it could be improved.

### **8.3 Growth of r/ebola**

r/ebola experienced two distinct periods of rapid growth, both of which correlate with specific events: the first began in July 2014, triggered by the news that an American humanitarian aid worker, Dr Kent Brantly, had contracted the disease in Africa and had been flown back to the U.S. for treatment. The subreddit expanded from 29 subscribers on 26<sup>th</sup> July 2014, the day on which Brantly was diagnosed, to more than 2,600 by the end of September 2014 (during this period the outbreak was also officially declared a PHEIC by WHO, on 8<sup>th</sup> August 2014). It gained another 10,000 subscribers over the first two weeks of October, following the first case of Ebola diagnosed in the U.S., on 30<sup>th</sup> September 2014 – a Liberian man who had contracted the disease in Africa and flown to visit family in Dallas before falling ill. Between 1<sup>st</sup>-15<sup>th</sup>

October, r/ebola had on average just under 165,000 page views a day, reaching a peak of 565,280 on 15<sup>th</sup> October, compared with an average of just 22,760 page views a day in the last two weeks of September. By November 2014, r/ebola had more than 14,000 subscribers. In contrast, r/cancer has not reached 10,000 subscribers in the nine years since it was set up (its subscriber numbers stood at 9,888 on 30<sup>th</sup> April 2017), and r/diabetes took just under seven years to achieve the same number, despite many more Americans living with those conditions that were ever at risk from Ebola.

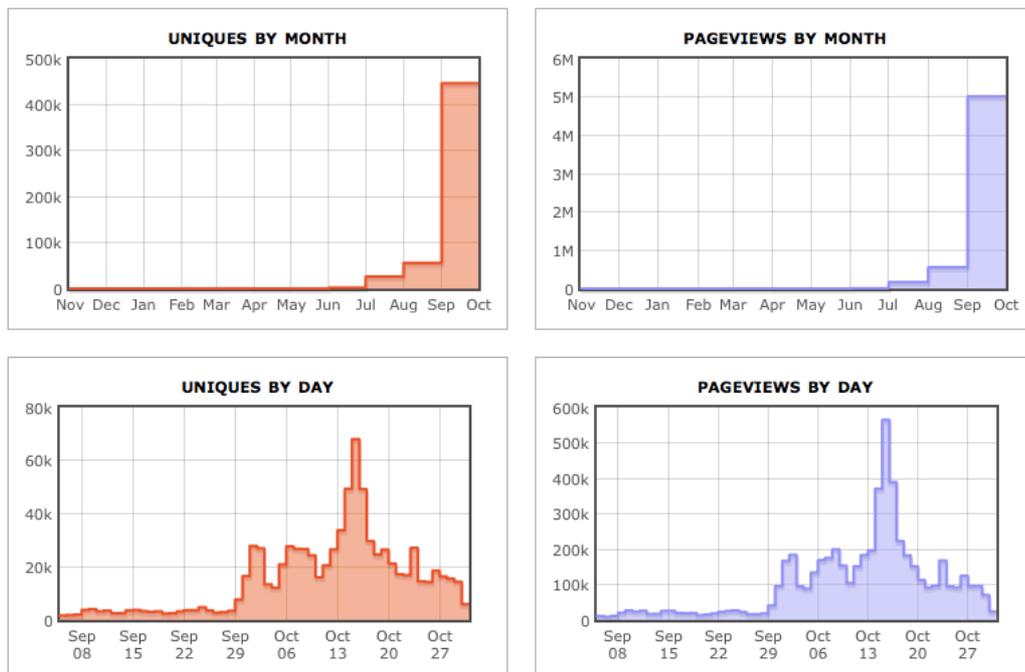


FIG 56: Page views by day for r/ebola from September-October 2014. The spikes in traffic coincide with the first diagnosis of Ebola on soil (30<sup>th</sup> September), the first recorded onward transmission on European soil (6<sup>th</sup> October) and news stories surrounding a nurse who was infected, and who was supposed to be in quarantine, taking a flight from Cleveland to Dallas (15<sup>th</sup>-16<sup>th</sup> October). The events that cause the spikes correlate with a possible perceived shift from the 'Far at Risk' to 'Near at Risk' categories', identified in Chapter 5, for reddit's predominantly U.S. and Europe-based user base.

## /r/ebola metrics (Ebola)

News, Information and Discussions about the Ebola virus.

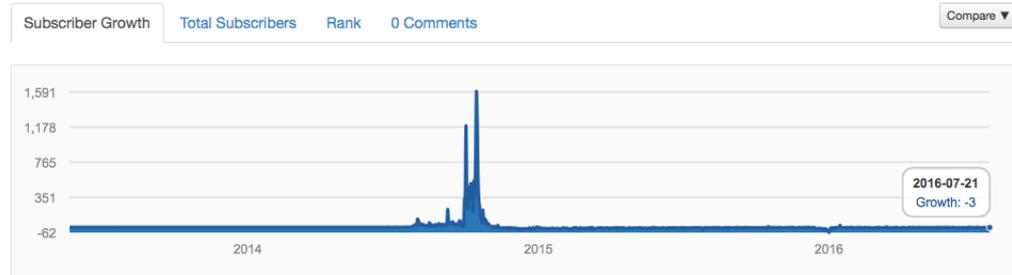


FIG 57: Timeline for r/ebola's subscriber growth, showing the spikes around October 2014, when a handful of U.S. and European citizens were being treated for Ebola. Since mid-November 2014, by which time it was clear that the cases in the U.S. would result in further spread and that the outbreak was under control in Africa, there has been a net loss in subscribers each day.



FIG 58: Spikes in activity on the r/zika subreddit have been less pronounced, but there was a flurry of activity in Jan-Feb 2016 when stories of the outbreak were prominent in the international media (the outbreak was declared a PHEIC on 6th Feb 2016) and around the 2016 Olympic and Paralympic Games in Brazil, where most of the cases have been recorded, in August 2016.

This provides some valuable insight into what exactly drives public interest in a PHEIC. In the case of Ebola, the ‘trigger’ for both upkicks in subscribers was an increased ‘nearness’ of the disease to reddit’s largely U.S.-based users, which may indicate that they perceived themselves to be moving from the ‘Far at Risk’ to ‘Near at Risk’ categories identified in Phase I of the study, and were actively seeking out information from a variety of sources (Cole and Watkins, 2015), which a news aggregator site such as reddit is well placed to deliver, making it an attractive destination for new information seekers.

This rapid rate of growth, and the potential of specific news events to trigger rapid increases in the number of subscribers and volume of traffic has interesting implications for platforms used for public health emergencies. Chapter 7 has shown that an important influence in quality assurance on reddit is the experience of the moderator, and their ability to identify and recruit good new moderators from the existing user base (a process referred to on reddit as ‘stocking’) based on users’ previous posting history and activity.

Such experience is best built up gradually over time, a luxury that may not be afforded in a PHEIC, though it is worth noting that signs a PHEIC is on its way, such as international media interest, generally predate the formal declaration by some months. There was eight months between the index case of Ebola and the formal declaration of a PHEIC, for example. The pandemic phase descriptions and alert system used by WHO are also designed to give countries an early warning of emerging diseases so that they can take action before a pandemic is formally declared.

A public health organisation may well have time to plan ahead if it is able to anticipate not only how the disease outbreak will develop, but also the specific events that will trigger media and public interest, enabling strong development of appropriate messages to reassure the public as well as on infection control.

## 8.4 Genesis of r/ebola

The genesis of r/ebola gives some insight into how the process has worked during a previous PHEIC and may work in future ones. r/ebola did not start out as a PHEIC-specific subreddit but was created on 19<sup>th</sup> May 2013, seven months before the index case in the 2014-16 West African Ebola outbreak. It was set up by u/jestersdoor, who had been a regular reddit user for more than two years prior to that, posting across entertainment-themed subreddits on music and gaming. He says he had no real intention of running an ebola-themed discussion or of moderating a forum, and claimed in interview that he created it, “just for the hell of it one day, when I was bored”. Despite this, the modmail logs for r/ebola show that he has maintained an active role in the subreddit, participating in 47 separate conversations during the Ebola outbreak period and has had an active role in configuring stylesheets and contributing to discussions on how the moderation team should deal with problematic users. The other r/ebola moderators I interviewed spoke of him positively, both because of his willingness to allow others who were prepared to develop r/ebola to do so, and for staying involved – his ‘hands-off’ management style was seen to ensure that none of the other moderators became ‘little dictators’ [RM008]. He could step into internal disputes between moderators from a relatively detached position, providing a genuinely independent oversight. r/starwars moderators [RM015, RM017] also spoke of a similar relationship with their original, but now less active, lead moderator.

At the time the disease was officially recognised by the WHO as Ebola, on 25<sup>th</sup> March 2014, however, r/ebola had just four subscribers and was doing nothing to promote itself. During the very early stages of the outbreak, there was effectively no dedicated Ebola subreddit but, as the disease spread, there was considerable discussion of it in other subreddits, particularly r/news, r/worldnews and r/science.

The two redditors who were to become the most active moderators in r/ebola ‘met’ in r/worldnews, “about March or April 2014. Both of us were posting and commenting on Ebola-related articles,” [RM008]. In their opinion, the information available was compromised by, “misinformed/misleading media articles [and] certain posters also had an unproductive stance. There were some that would always focus on the worst-case scenario and spin information to that effect ... and wildly inaccurate predictions of total cases” [RM010]. They wanted to find a forum more specific to Ebola that could be used for “reputable news and balanced discussion” [RM010], and moved across into the already existing, but at that time low-activity, r/ebola, where they started posting regularly. Soon after, they were contacted by the original moderator, who offered them moderator status. The invitations they were sent – and accepted – represent the earliest communications in the r/ebola modmail logs.

Then “things picked up in a hurry” [RM008] at the end of July/beginning of August 2014, coinciding with the first spike in subscriber growth. This caused some problems for the relatively inexperienced moderators who, while they were familiar with the subject and with reddit’s structure and norms, had less knowledge of the technical aspects of the moderation process or of the likely challenges of moderating a disease outbreak forum. The role proved to be extremely time consuming. At the height of the outbreak, the most active moderator reported in interview spending eight hours a day or more working on the moderation tasks (including posting new information). Of the total modmail correspondence logged on r/ebola modmail between July 2013 and July 2016, 56% was sent in one month alone – October 2014 – and 89% was sent between September and November 2014.

By the beginning of August 2014, the subreddit was growing rapidly – with further and more rapid growth anticipated, particularly if predictions of the number of new cases proved correct. At the time, even legitimate sources such

as WHO and CDC were warning of possible case numbers worldwide between 100,000 and 1 million (the actual figure stood at 28,616 cases and 11,310 deaths at the point the PHEIC was lifted on 29<sup>th</sup> March 2016). Along with the increased number of posts, however, came numerous offers of help. As interest in, and traffic to, the subreddit increased, potential new moderators – with “advanced moderating and operational skills” [RM008], came forward and offered their services. These people had a combination of skills, including experience of reddit, subject matter expertise, experience of moderating and in some cases, very specific knowledge of very particular areas of moderating, such as not only how to set up automoderator, but also of the type of language usually associated with negative behaviour. They knew of redditors who regularly exhibited racist or troll-like behaviour and of general miscreants who caused trouble across other subreddits.

Moderators who are willing to join the team at this stage may have very different motivations to those who set up the subreddit; rather than wanting to create something they feel does not already exist, largely because they want it to exist for their own use, they may be more interested in enabling what already exists, and which they see as valuable to the reddit, continuing. “They see its worth and [step] in to help”, [RM008] – as much with the mundane, housekeeping tasks as with content management.

In the middle of August 2014, the existing moderation team reached out through Pro-MED mail (an alert system operated by the Program for Monitoring Emerging Diseases) to ask for volunteer moderators with infectious disease or specific Ebola knowledge, as most of the existing moderation team were “structural moderators” [RM008]. At this point, the sub was receiving 60,000 page views per month. The moderation team also discussed applying for ‘subreddit of the day’ with the intention of deliberately intending to push traffic up to 10,000 page views a day. However, the

increased volume of traffic meant they were also beginning to attract racist comments and other negative content, such as an increasing number of links to sensationalist media and self-posts calling for travel bans and quarantines to be imposed. There was some disagreement amongst the moderation team about whether a deliberate attempt to increase traffic further should be pursued: not all felt the community would get better with a huge influx of people. The moderators with experience of larger forums and/or forums which have seen rapid increases in traffic [RM015, RM030] both agreed that this had not always been advantageous. The lead moderator did however eventually apply for, and was granted, the extra exposure with no objection from the moderator who had highlighted the potential downfalls, and support from the rest of the team.

As traffic grew, the moderators also started to discuss ways to identify the good information and ‘filter’ it from the less good (which required extensive understanding of how to use the Automoderator tool), to cut down on the amount of time needed for human intervention. They also attempted to bring in more subject matter expertise. A call-out went to Pro-Med, and later to r/virology and r/epidemiology, but had had little response. This may have been due to lack of time to follow-up – the moderators interviewed suggested that in a future event, having a ‘liaison officer’ whose role was to initiate and maintain contact with other groups, internal or external to reddit, might prove useful. Librarians, or those with experience of how to index and order large amounts of information was another skill that was identified as potentially being useful – interesting to note in light of Pierre Lévy’s research interests into semantic ordering systems for the masses of information on the world wide web. The moderators also discussed other strategic directions for the site, such as assigning flair to posts, to identify more reliable sources such as WHO and CDC, and to flag ‘media’, ‘speculative’ etc (flairs for CDC, WHO, media, speculative and others were phased in from mid-August onwards). A filtering

system based on that used by r/technology was also discussed and some work was carried out to develop one, but it was never completed due to time constraints. During this period, the moderators also sought advice from other forums, particularly r/worldnews, from which they had emerged.

If any of you want to help with restating the question, please do so. If any know of anyone with direct hands on knowledge on this matter, put me in contact with them.

Hope there's an establish method in place (which has proven successful) we can modify for our own purposes. One similar to the filter system on /r/technology/ might work as a template.

Possible Filters (for consideration only, not putting these forth as usable):

- Vetted as factual, posts Mods have verified as being accurate and truthful.
- Speculation, conspiracy theorists, What-Ifers and the like.
- From the Hot Zone (need to change to a phrase less sensational) reports from those living in countries where the disease is epidemic.
- Pure Bullshit (need to change to a phrase less offensive)
- Self-Interest Only, the ones focused on the individuals future action. ie is it safe for me to travel to .....
- Report from News Media source - companies/corporations who make their money selling advertising.
- Report from governmental agency - WHO/MSF/CDC/MOH, etc.
- Haters - racists, nationalist and the 'who needs Africans' crowd.
- <any label you think is not here, but needs to be>

## Technology

subscribe

5,053,722 readers  
● 5,219 users here now

Join the live chat on IRC

Browse categories:

Security	Networking
Hardware	Software
Robotics	Business
Politics	Biotech
Transport	Space
Energy	Wireless
Nanotech	AI

FIG 59 (above): An entry from r/ebola modmail logs showing discussion on how to assign filters to posts, citing the system in use on another subreddit, r/technology, as a potential model to follow.

FIG 60 (left): The system on r/technology (left), a large subreddit with millions of users, assigns flair and then allows users to filter the submissions to focus only on certain categories.

During the period of most rapid growth (August-October 2014), most exchanges via moderator mail were still predominantly to do with technical administration, such as setting spam filters, along with discussion around what the tone and positioning of the sub should be. A challenge was how to deal with conflicting information from official sources, which made it difficult to determine what the facts were; speculative media reports tended to focus on worst-case scenarios and death predictions. Moderators also discussed whether to allow posts linking to personal blogs and how to manage the anticipated increase in traffic. All these are areas where moderator experience is extremely useful as it allows experienced moderators to advise on whether, for example, an increase in subscribers and page views is likely to see a corresponding increase in posts and comments, or whether different elements of participation increase at different rates. Importantly, however, r/ebola was defining itself as it grew – unlike on subreddits that experience more gradual growth, on r/ebola setting up the site and growing it happened simultaneously, over a remarkably short period. Evident from the moderator mail logs is that as well as undertaking the routine tasks on the subreddit, during this period the moderators were also having to keep on top of fast-moving news and the latest scientific developments themselves, to make decisions about what did or did not constitute good information. In the chronic condition subreddits, the moderators tend to have lived with their conditions for years, they know the questions newcomers are likely to ask, the type of negative behaviour the forum might expect and have a ready-made community, whose members they are familiar with, from which to recruit additional help when needed. In r/ebola, everyone was a beginner in the topic of an Ebola outbreak on this scale, even microbiologists, doctors and epidemiologists, a situation akin to that in which The Delphi Method emerged during the Cold War, when the impact of the anticipated situation could be based on speculation only rather than experience. In addition, when spam filters and automoderator are being modified and beta-tested almost daily, it is inevitable that an additional

number of legitimate posts may get ‘stuck’ until the filters are properly defined, requiring already time-pushed moderators to manually approve them and to reply to confused users wondering where their post has gone.

The importance of providing accurate, factual data at this stage of the outbreak – the ‘Far at Risk’ phase – also prompted a need to provide such information in a format that could be read and understood easily. Data tables of cases, some compiled by the moderation team and some enabled by link posts to organisations such as WHO and CDC, became important and proved popular with the community. A self-post containing a ‘Comprehensive List of Confirmed and Reported Cases Outside West Africa’, posted (and compiled) by approved poster u/briangiles on 4<sup>th</sup> August 2014, and updated regularly until 6<sup>th</sup> November 2014, received 650 upvotes, for example, and is the 6<sup>th</sup> most popular post ever made on the subreddit.

This period also began to see r/ebola attracting deliberately negative behaviour, including suspected coordinated downvoting, by conspiracy theorists or posters who would twist facts or support misinformation – most commonly the suggestion that the Ebola virus could, or had, become airborne – and overt racists from white supremacist subreddits including r/coontown. There was also evidence of upvoting from username accounts suspected of being a ‘sockpuppet army’ – multiple accounts used by a single individual to promote their own agenda. The influence of such behaviour on reddit is somewhat limited due to the ability of each account to only have one registered vote: a single individual would need to have dozens of accounts to have any serious influence on the hotscore – for instance, to elevate a post to the Front Page – but they can be troublesome nonetheless. Decisions had to be made about whether, and how, to ban such users, or how to configure automoderator to stop others like them if deemed necessary (banning a user does not stop them subscribing to and reading the subreddit, only from

contributing to it. A banned user can still be educated with good information, but is not able to spread misinformation or be abusive to others).

New users also proved to be a challenge. A well-used tactic on reddit is to set automoderator configurations to ‘catch’ posts from user accounts less than one day old, both because these tend to be ‘throwaway accounts’, which may have been created simply to post abusive comments (Leavitt, 2015), and also because even when new users are genuine, they may not understand the rules and norms of the forum, and post inappropriate material, material that has been posted many times before, or material that does not conform to the forum’s chosen position. New users within r/ebola may not be new to reddit, however, meaning that this filter will not work: established redditors coming in from other subreddits, such as conspiracy sites or racist forums, would still get through. While this filter can be set to enable only subscribers to post, or only approved posters’ submissions to be visible immediately, doing this is complicated and time-consuming.

During the second half of August, the moderators addressed the increasing volume of information of questionable quality by the addition of flairs (see FIG 61). These were introduced to help users make a value judgement on the likely validity of the information based on its source, and by expanding the list of links to reliable sources of information in the front page’s sidebar.

By mid-September 2014, most of the Automoderator settings, spam filters, flairs and policy discussions seem to settle down, and most of discourse in the modmail forum consists of replies to users curious as to why their posts have been caught by spam filters, and banning users or discussing the bans. Around this time, a decision was also made to ban all solicitations for donations to ‘help fight Ebola’ as the moderators did not have the time or the ability to determine which ones were legitimate and which were not.

Flair description, purpose and usage	
Flair Elements	Descriptions, Purpose & Usage
<b>CDC</b>	Posts linked to <a href="#">Centers for Disease Control and Prevention</a> articles
<b>Discussion</b>	<i>Discussion</i> posts are a form of <i>self</i> posts used to promote discussion of Ebola related topics
<b>Economic</b>	Posts linked to articles with the primarily focus being on economic factors
<b>EduSIG</b>	Posts linked to Educational based special interest groups such as: <a href="#">CIDRAP</a> & <a href="#">UCANR</a>
<b>GOV</b>	Posts linked to government agencies other than health departments
<b>HDR</b>	Posts linked to National and International Health Department reports
<b>LANCET</b>	<a href="#">The Lancet</a> medial journal publications
<b>MEDIA</b>	Posts linked to any local, national or international Media (News) publishers
<b>MSF</b>	Posts linked to <a href="#">Doctors without Borders/Médecins Sans Frontières</a> articles
<b>MSTAgg</b>	Posts from Medical/Scientific/Technical Aggregator websites
<b>MSTBlog</b>	Posts from Medical/Scientific/Technical Blog websites
<b>NEJM</b>	Posts from <a href="#">New England Journal of Medicine</a> publications
<b>NIH</b>	Posts from <a href="#">National Institute of Health</a> and associated 'NI' agencies
<b>Other</b>	The <i>Other</i> Flair is for posts which do not match any of the other flairs.
<b>Political</b>	Posts linked to articles with the primarily focus being on political factors
<b>Science &amp; Medicine</b>	Posts linked to Scientific & Medical website that are not blogs
<b>Self</b>	Self posts are used to make statements specifically related to Ebola
<b>Sticky</b>	The <i>Sticky</i> post is used exclusively by the Ebola Moderators on posts being highlighted by placing them at the top of the list of postings.*
<b>Unconfirmed</b>	Posts containing information which has not been substantiated
<b>Video</b>	Posts to video materials related to Ebola
<b>VizData</b>	Posts to Data Generated visual representations
<b>WHO</b>	Post linked to publications from the World Health Organization
-----	
<b>Africa</b>	The 'Continental' flair tags: Africa, Asia &
<b>Asia &amp; Oceania</b>	Oceania, Europe, North and South America
<b>Europe</b>	are 'catch-alls' tags automatically generated
<b>North America</b>	by the subReddit and may or may not be related
<b>South America</b>	to the continent the post references in its link

FIG 61: *List of the flairs available on r/ebola as of 27<sup>th</sup> July 2016.*

Activity picked up again in October, with confirmation of a case of Ebola in Dallas, the first to be diagnosed on U.S. soil. The r/ebola moderators made a second attempt to bolster their moderation team, but even with Automoderator well-configured, the quality of the posts was considered to drop [RM007]. Interestingly, although the influx of new users was proving problematic, the moderation team still considered the number of subscribers – 1,500 on 10<sup>th</sup> September, to be disappointingly low and discussed measures to increase it.

This kicked off the discussions of whether to apply for ‘subreddit of the day’ – awarded at the Admins’ discretion – for which r/ebola was selected on 2<sup>nd</sup> October 2014, at a point when it had just over 3,000 subscribers. The Admins warned them to expect more traffic because of this, and traffic did indeed increase over the following two weeks (there was a tenfold increase in page views from September to October), with an initial drop in quality similar to that witnessed with the first leap, which was recognised by users and discussed on the subreddit itself:

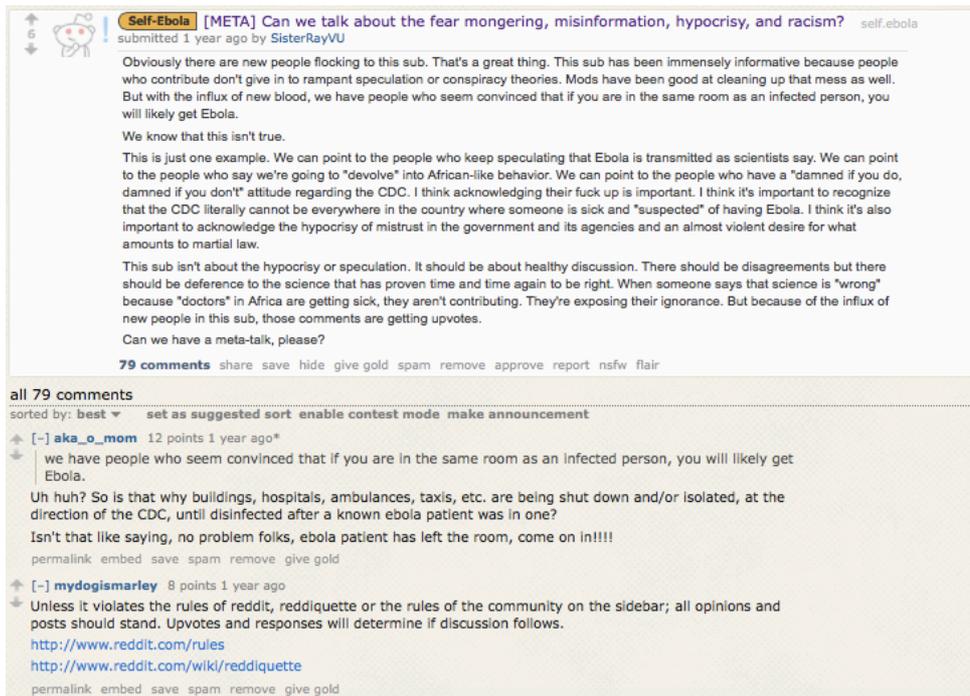


FIG 62: A discussion started on r/ebola on 2<sup>nd</sup> October 2014 by a user who had become frustrated with what they perceived as a decline in the quality of content over the past few days.

While the moderators did their best to cut down self-posts and repetitive topics, they found it hard to agree what to do in the short-term other than mass bans and deletions, which they were loath to implement. On 4<sup>th</sup> October, they decided to try to develop an FAQ, which could become a permanent feature and provide some relevant facts and figures on Ebola.

There was also a second call for moderators and discussions of (again) trying to attract moderators with subject matter expertise, potentially by deliberate targeting of redditors with science or medical user flair, who were regularly posting about Ebola on other science-related forums such as r/science.

Moderator motivation to keep discussions factual and high quality was particularly tested during this period, with users messaging the moderation

team to suggest/demand the subreddit give more prominence to the issue of Ebola ‘escaping’ out of Africa to Spain and the U.S. The moderators maintained a measured and level-headed approach, replying patiently to users who sent messages and doing their best to temper the more speculative media stories and self-posts appearing on the site. Some reddit users will post news items on topics they think will prove popular as a way of attracting upvotes to boost their karma (known colloquially as ‘whoring for karma’) meaning that popular topics can be a vicious circle, attracting more interest because of the popularity, rather than the topic under discussion itself.

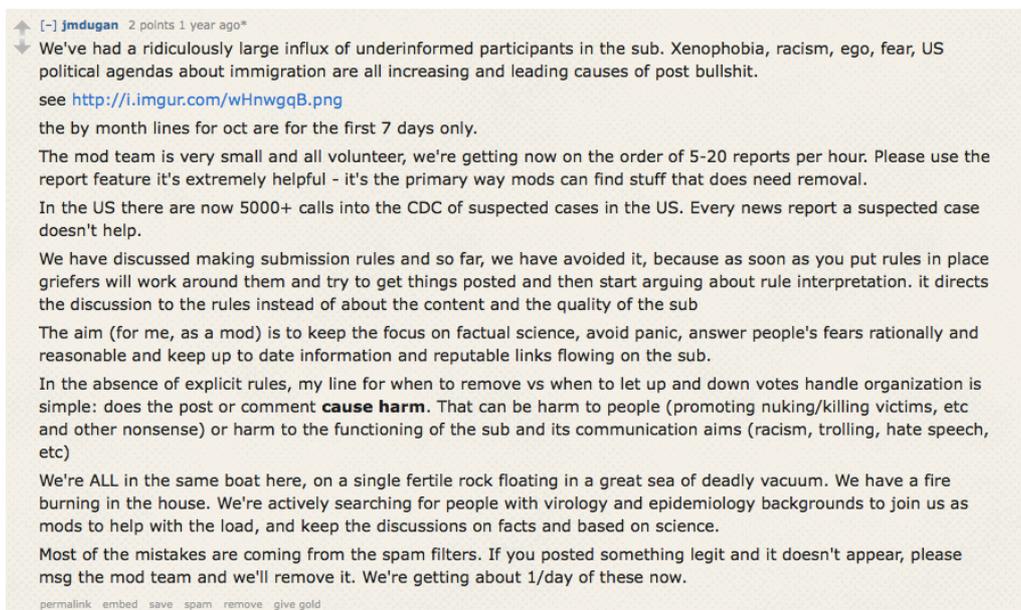


FIG 63: *The moderators’ reply to a discussion on the removal of posts. In a later child comment, they also explain that the moderation team have undertaken 349 moderator actions between them in the previous 24 hours alone, stressing the pressures they were under at this time.*

There was still site admin to maintain during this period, including deciding on a stylised reddit alien (logo) for the subreddit, with a number of designs considered before settling on one, and by mid-October it was still felt that more mods were needed, resulting in a further call-out for help. Three new

moderators were added to the team by 16<sup>th</sup> October, when traffic to the subreddit was its busiest. The modmail forum was also hosting conversations on policy decisions, such as to what extent the subreddit should support the CDC and what level of criticism of it should be tolerated, as well as just trying to keep the subreddit running.

What was happening highlights the challenges associated with needing to expand a moderator team quickly, and of bringing in less experienced moderators at a time when the moderation team is most busy and most under stress. Three new moderators were added on 16<sup>th</sup> October 2014: one was an experienced reddit moderator who had worked across several other subreddits, one was a regular poster who was identified as a likely good content editor (she had previous experience of moderating large online forums, but not on reddit) and one was a graduate student with subject-matter expertise but less experience of moderation or, in fact, of reddit in general. Over the following two days, the inexperienced moderator had a negative impact on group dynamics, constantly suggesting changes to site design, to Automoderator configuration and to policy, such as whether to allow fundraising to be promoted on the site. He sometimes implemented these changes without consulting the moderation team, or the moderator(s) primarily responsible for specific actions, such as site design. This resulted in some disagreement within the moderation team. Some of the existing moderators saw his actions as heavy-handed, and beyond what the forum needed on its busiest day(s): they saw the priority as simply dealing with traffic and trying to keep on top of the flood of interest and the volume of posts – often inappropriate – being pushed to the site by the first diagnosis of Ebola on U.S. soil, and the onward transmission from the patient to two healthcare workers who had treated him. Others felt the fact that “he jumped in and got stuff done” was beneficial. One of the other two new moderators, the one with the most previous experience of reddit, tried to explain to the disruptive moderator that the “busiest two days in

the subreddit's history" was not the time to be suggesting major changes to structure and policy – and that while his ideas might be good, his approach was not. He was removed from the moderation team on 17<sup>th</sup> October by the most active moderator, less than two days after being invited to join. The discussions on modmail show that the decision was not unanimous between the other moderators, but while any one could have re-invited him, none did.

The event highlights the specific challenge of needing to reach out to subject matter experts who may not understand the process of how to moderate or be part of a moderator team, during times of stress for a PHEIC subreddit.

Following the disruptive moderator's departure, the moderation team did consider making some of the changes he had suggested – in fact, soon after his removal the moderator who had been his most vocal opponent and had the most say in his removal, instigated a discussion of which of his suggestions might be incorporated as, "some of his ideas are worth considering".

For the moderator team, however, the experience presented a dilemma: subject experts were needed, but integrating them into the moderation team was challenging. This was particularly problematic as the upsurge in interest had resulted in several valid and valuable posts from genuine new accounts, leading the moderation team to disable the spam filter which removed posts from accounts less than one day old – usually used to prevent 'throwaway account' trolling – but this also increased the manual moderation burden.

Instead, an alternate strategy was suggested: reach out to subject matter experts (in science, African politics, anthropology, military – as the military response was gearing up – and medics and survivors from the 'front lines') and invite them to take part in discussions on the site, but not invite them to moderate. The subreddit reached out to r/virology and r/epidemiology to ask its flared users to "check in now and again to reinforce the science and facts",

particularly to answer questions that required a scientifically accurate answer. The need to reinforce the moderation team was also acknowledged, and the most experienced of the two remaining new moderators reached out to other experienced moderators she had worked with on other subreddits and invited them to join. Two accepted and, while not subject matter experts with regards to Ebola or public health, they were experienced reddit moderators who both proved to be valuable additions to the team.

Another challenge raised by the increased interest in Ebola was that more and more of the posts being made to r/ebola were beginning to appear on the site-wide Front Page. In the opinion of the moderators, it was these posts that generally attracted most of the negative behaviour, in particular trolling, as this was most often seen in comments made in response to these Front Page posts. The solution was to ensure that such comments were closely monitored and removed, as well as banning the posters making them. This again was very labour intensive, however, but did have the advantage of identifying new terms regularly used by such posters that could be added to the Automoderator filter. The moderation team also reached out to some of r/ebola's regular posters to ask their opinions on how the sub should handle the challenges it was facing. One response that came was that regular users were having difficulty sorting the good information from the bad, as they did not have time to read and assess everything and so needed to be directed to the content more likely to be reliable; and that the increase in (over)speculative media articles could sometimes give the impression that r/ebola was a "panic sub". Some felt that the sub was over-tolerant of posters who appeared to be "apologist(s) for WHO and CDC", and that too many posters submitted content – particularly statistics and numbers – without providing reliable sources to back them up. There was also a feeling that self-posts from people who were overly worried about the threat of Ebola – to the world or to the U.S. – were being treated somewhat harshly. Such posts were generally removed, but not before the

poster had been shamed, demeaned and humiliated by more informed posters. Their fear may not be rational but trying to answer people's concerns, even if they were ill-informed and alarmist, was considered worthwhile as education about the virus would help them. There was a discussion about potentially giving more prominence, or permanence by way of a 'stickied post' – a post 'frozen' to the top of the subreddit front page whose position is not affected by the hotscore – to an FAQ originally posted by a user on 12<sup>th</sup> October. The team eventually settled on using the FAQ as a basis for a more thorough one, and invited the original poster to help shape and update it, with restricted moderator permissions that allowed him to work on the FAQ but not other aspects of the site.

A further issue was a promoted post (i.e. advertisement) that appeared at the top of r/ebola, over which the moderators had no control, which seemed to encourage over-cautious behaviour on stockpiling and social distancing.

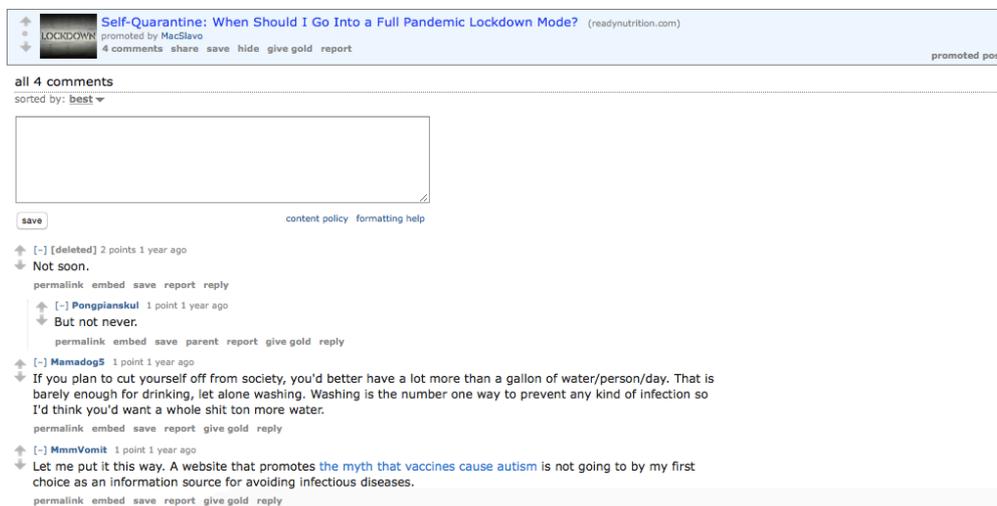


FIG 64: An advert that appeared on r/ebola at the height of the outbreak. The community could comment on the sponsored post (though not vote) – which it did so negatively – helping to dilute any influence the advert may have had by showing that the community gave it little credibility.

This advertisement appeared to promote the idea of Americans going into ‘self-quarantine’ and linked to a website that not only suggested U.S. citizens should start thinking about taking their children out of school and stock up on food and other essentials in case a ‘full lockdown’ was required, but also presented anti-vaccination sentiment and other views in line with Survivalist or ‘Prepper’ rhetoric – groups preparing to be self-sufficient in the wake of a perceived impending breakdown of society – which r/ebola had taken pains to filter out through the Automoderator settings. The community collectively reacted against it, with a series of comments that signposts its lack of credibility with the receiver community.

By 22<sup>nd</sup> October, activity on the site had quietened down enough (to 91,681 page views, from 565,280 on 16<sup>th</sup> October) that the moderation team could return to some discussions on site policy, such as whether and how to assign flair to users on the basis of qualifications (and if so, how to verify), expertise or good conduct; make amendments to the guidelines displayed on the front page regarding acceptable submissions and behaviour; and to begin more in-depth work on the proposed FAQ, the final version of which was posted at [www.reddit.com/r/ebola/wiki/index#wiki\\_frequently\\_asked\\_questions](http://www.reddit.com/r/ebola/wiki/index#wiki_frequently_asked_questions) on 24<sup>th</sup> October. Judging by the comments and voting it attracted, it was received favourably by the reddit community.

By November 2014, the ‘rush’ was over, and moderator actions returned to being more focussed around mundane ‘housekeeping duties’ – generally explaining to users why they had been banned and/or their posts had been removed by Automoderator, and approving legitimate posts that had been picked up overzealously by the filters. By the end of the month, one of the newer moderators brought in during the rush decided to step down, and this triggered some interesting discussions around centralisation of roles and tasks relating to the thread. At the height of the busiest period, moderators had often

been recruited due to specific expertise, to handle discrete aspects of the subreddit such as updating the spreadsheet of confirmed cases, suspected cases and deaths, or configuring Automoderator. As the rush subsided and some of those stepped away, it became important for the remaining core team to become more generalist, and to perhaps develop the skills required to maintain elements of the site its users had come to expect, rather than only those that were essential and which had been maintained before the rush.

From December 2014 to Jan 2015, the moderators continued with general activity. They also had to deal with the return of a particularly unpleasant racist element, but by this time they had the time and space to consider what to do. The discussion was particularly complex as one new poster who had appeared was well-known across reddit as a regular contributor to racist subreddits (and a moderator of one white supremacist forum) who made racist comments on others, though he appeared to be making reasoned and balanced submissions to r/ebola. Eventually a group decision was made to ban him, based on one moderator's previous experience that he often 'infiltrated' subreddits by posting balanced views, but soon after would encourage other known racists to also come along and descend into racist behaviour. This experience of the behaviour of specific redditors shows the immense value of experienced moderators who have worked across more than one forum. In general, however, activity between January 2015 and December 2015 was low, with two more moderators stepping down as the workload diminished. Most conversations during the entire year related to general housekeeping tasks. A random increase in porn-related spam in Jan-Feb 2016 was easily dealt with through Automoderator configuration but prompted a return to removing, rather than reporting, posts from new accounts as spam was now by far the predominant type of new account posts. A downturn in activity on the r/ebola modmail forum corresponds with a downturn in interest in the subreddit itself (there were 1,158 page views in January 2017, with the

average number dropping steadily over the previous year), and a more relaxed schedule on how often the WHO issued Ebola updates – and correspondingly how often r/ebola need to be updated to keep up. From this point on, the r/ebola moderation mail has been almost exclusively taken up with Automoderator notifications of removed posts – most of which have been pornography spam.

The period from July-December 2014 represented a remarkable learning curve. During those six months, people with no experience of moderating a large subreddit (or in most cases of experiencing a subreddit at all), came together to create a forum dedicated to providing information on a fast-moving outbreak on which few people would have any real experience. They not only weathered the storm, but built up a valuable catalogue of lessons identified and in most cases learned that could be stored up for future use. And future use was not far away.

## 8.5 From r/ebola to r/zika

In Feb 2016, a message was sent to the r/ebola moderators from the moderator of r/zikavirus, asking for their help in moderating another subreddit dedicated to an emerging disease.

**[–]from [deleted] to /r/ebola sent 5 months ago**

Hello!

Due to the abandonment of [/r/zikavirus](#), I have taken control to ensure it didn't enter the hands of someone who may have had an agenda of some kind. However, I have very limited experience regarding the more technical side of moderation, so more help would be appreciated.

I have asked you because this seems to be a very professionally moderated medical subreddit, which is exactly what I am hoping to build [/r/zikavirus](#) into.

If any of you have the time to spare, please send me a message.

Thank you for your time.

[/u/TheLazySocialist](#)

[permalinks](#) [spam](#) [remove](#) [report](#) [block](#) [usermute](#) [usermark](#) [unreadreply](#)

FIG 65: *Message from the moderator of r/zika to the r/ebola moderation team.*

The moderator was interested in the subject, but had little experience of moderation or the technical side of moderation, and so was asking the r/ebola moderators for their expertise in this field. Around the same time, the r/ebola core moderators [RM007 and RM008] also contacted the moderators of another Zika subreddit – r/zika – that had recently been set up, offering help with the moderation queue and moderation tools such as Botwatchman and the advanced Reddit Enhancement Suite (RES) (of which the current moderators had been unaware). They also helped with the creation of additional subreddits for different languages, including Portugese (the national language of Brazil, where most of the case were being recorded), French and Spanish. This had an added benefit of introducing the r/ebola moderators to a new moderator with technical expertise in CSS programming who, after initially helping with the r/zika language subreddits, also helped with the CSS programming on r/ebola.

Regarding the two parallel Zika virus threads, r/zika\_virus and r/zika, the r/ebola moderators eventually settled down into r/zika (largely due to what appears to be a clash of personalities between some of the individuals involved, though the one on the r/zika\_virus side has since deleted their reddit account entirely; the subreddit now lists three moderators, none of whom were involved at the start and cannot (or do not want to) shed any further light on what happened. The subreddits share some posters (and posts) and the content and tone of the two is largely indistinguishable: they coexist in relative harmony. r/zika has more subscribers (449 on 22 Feb 2017, while r/zikavirus has 74) but neither has attracted the attention that r/ebola did.

The moderators coming across from r/ebola to r/zika reached out to good posters from both r/ebola and r/worldnews and asked them to subscribe, helping to ensure that early posts made to the sub were likely to be of high quality. This also included reach-outs to regular r/ebola posters known to be fluent in Spanish, to help with the Spanish language version and translations

of Zika articles written in Spanish. The requirement for additional language knowledge was another learning experience from r/ebola, which had anticipated a potential need for subreddits in local African languages and major world languages if the virus had spread out of Africa. By 4<sup>th</sup> February 2016, less than two weeks after r/zika was set up (on 27<sup>th</sup> January), its Automoderator was configured with the benefit of r/ebola's hindsight and was filtering out posts relating to the conspiracy that genetically-modified mosquitos were somehow responsible for the outbreak, that the outbreak was linked to Roswell Conspiracy (that an alien landing in the USA in the 1950s was covered up by the Government), highly speculative media items about unverified reported cases, porn spam, and several unhelpful or misinformed comments that had been posted from brand new accounts. A moderator-only discussion forum, r/zika\_mods had been set up to enable better internal communication, and by 15<sup>th</sup> April, when traffic was still slow, posts were already able to be flared with various indicators of potential quality (or not).

As of mid-2017, the Zika outbreak has not attracted the same level of attention as Ebola, either in the international media or on reddit. There have been no sensational 'events' equivalent to Thomas Eric Duncan's diagnosis that might push large numbers of ill-informed and inexperienced 'newbies' to the r/zika or r/zikavirus subreddits, nor make them attractive targets for trolls or spammers. But, at the end of July 2016, when r/zika had less than 300 subscribers and just 500 posts had been made in its history, it had a strong moderation team comprising subject matter experts (including a Professor of Public Health), experienced reddit moderators, CSS programmers, and speakers of the first languages spoken in the countries most affected. Automoderator has been extensively configured to filter out conspiracy theories, racism, sexual and religious abuse. It had been set to automatically approve posts from more reputable national and international media sources including the BBC, New York Times and Reuters while blacklisting those

from more sensationalist media such as the Daily Mail and known conspiracy sites. It will report/remove posts associated with behaviour known to be consistent with low quality submissions, such as words in all capitals, new accounts, posts that use too many question marks (????), profanity and from posters with negative karma scores. It has banned known troublemakers; provided a series of flairs to enable users to make quick judgement calls on the content of links; identified 19 approved submitters and has set policy such as not supporting fundraising efforts and not accepting memes and jokes. Clear rules for contributing are set out on its homepage, along with links to all the main international health organisations. A possible upsurge in traffic around the 2016 Olympic and Paralympic Games in August 2016 (which did see a moderate increase in activity) was anticipated and prepared for. r/ebola did not achieve this level of organisation and preparedness until late October 2014, when it was on the downsurge of its subscriber growth peak, had more than 10,000 subscribers and had hit half a million page views a month.

While r/zika may never attract the interest r/ebola has, should that day come the system – both human and electronic – is in place to cope with it, and has emerged from the collective knowledge, experience and community of reddit to create something unique.

## **8.6 When do surges occur?**

Documenting the history of r/ebola and r/zika gives some insight into when surges in interest in the subreddits occur, and what events trigger them.

Trigger events can be external or internal to reddit: in the case of r/ebola, the external triggers included reports that American healthworkers with Ebola had been flown back to the U.S. for treatment, and that a case of Ebola had been diagnosed on U.S. soil. Both events cross the boundary from ‘Far at Risk’ to (at least perceived) ‘Near at Risk’ identified in Phase I of the study. These

events drew far more subscribers, comments and submissions to r/ebola than, for example, the announcement by the World Health Organization that the outbreak had officially been declared a PHEIC on 8<sup>th</sup> August 2014.

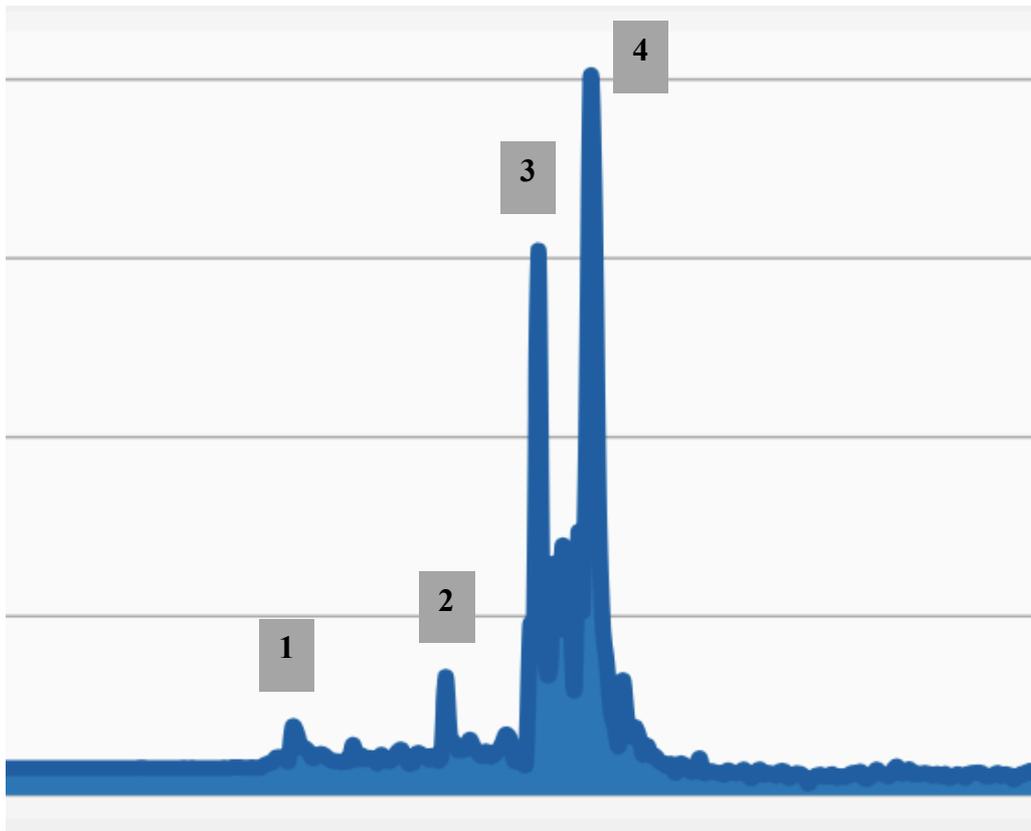


FIG 66: *Pattern of new subscriber growth on r/ebola. The peaks correspond, from left to right, to [1] First U.S. healthworker (Dr Kent Brantly) infected, [2] WHO warns of likely exponential growth; Bill Gates pledges major financial donation; infected U.S. healthcare workers flown back to U.S. [3] Thomas Eric Duncan diagnosed with Ebola on U.S. soil after travelling back from Liberia; [4], Texan nurse makes an internal flight before being diagnosed with Ebola. There is no notable spike in traffic for either the official recognition of an Ebola outbreak by WHO on 25<sup>th</sup> March 2014, or the declaration of a PHEIC on 8<sup>th</sup> August 2014.*

This makes the exact timing of likely surges difficult to anticipate in the case of public health emergency subreddits – it appears that it is not the emergence of the disease itself that drives interest, but the apparent proximity of the cases to the system’s users. Other subreddits that have experienced similar rapid increases in interest tend to find that these happen around planned and anticipated events – such as the release of a new album on a music subreddit, or a new movie or book on an entertainment subreddit. *r/starwars* sees huge increases in traffic each time a trailer for a new movie comes out, as well as for the new movies themselves, which provides ample time to bolster the moderation team and to forward plan: “I was pushing to add more moderators pretty early in the process, even six months out, so most people had chance to do some things before we got hit that bad. Most of the moderator team that existed then ... most of them don’t moderate anything that large, so they’re not really familiar with that type of traffic, and the sort of misbehaviour that comes with it” [RM015].

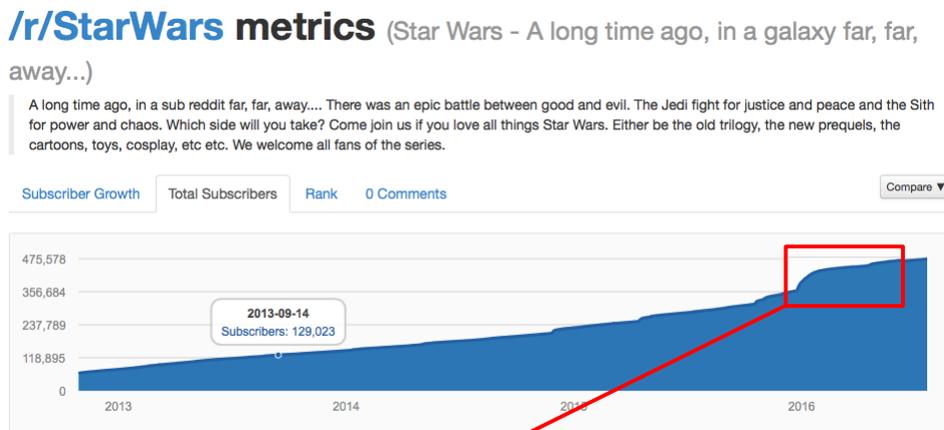
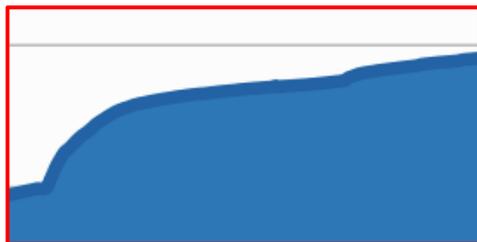


FIG 67: *Subscriber growth on r/starwars, which experienced a spike in subscriber numbers around the release of a new Star Wars movie in December 2015 (from www.redditmetrics.com).*



Moderators interviewed also reported regular surges occurring after AMAs – ‘Ask Me Anything’ question and answer sessions with experts. “We got one good surge following an AMA from a scientist who did research [on this condition] that made the Front Page” [RM001]; “when news related to a sub happens...” [RM008]; and when “someone dies” [RM015]. Moderators suggested that it can be useful to have FAQs ready when a surge is anticipated so that the same questions aren’t asked more than once, and to ask users in advance not to post links to the same news items other have already posted.

## **8.7 Discussion and conclusions**

Surges can have several impacts on subreddits, which can be challenging for both moderators and users if not handled carefully. Moderators described increased interest as “a double-edged sword, having hundreds of people who don’t know the rules and who go crazy and stuff” [RM004], which could require, “a fair bit more time to keep things under control” [RM007]. In particular, the same information can often be posted multiple times; “you’ll have three or four hundred people posting [links to the same movie trailer] and we don’t... I mean, we only need it once (laughs)” [RM015]. While Automoderator can be quickly configured by an experienced moderator to filter out multiples – by, for example, choosing a word that is appearing in every post title – this often needs to be done in real time, and requires a moderator comfortable with setting configurations to be on hand and ready.

Events internal to reddit that can create a surge in interest include a post (and by extension the subreddit to which it belongs) appearing on the Front Page, and the internal promotion of the subreddit through being chosen as a ‘Subreddit of the Day’ or for an advertisement within the site. As a window into the rest of reddit, the Front Page is a major driver for increasing participation and visibility. Posts that get to the Front Page solicit a huge

amount of interest, which can be considerably above the number of subscribers its subreddit has and can draw the attention of other reddit users to its existence. This has both positive and negative consequences. “As soon as something hits the top 25 on the reddit Front Page, you suddenly just see an enormous jump in not just the number of comments, but the number of people who are just looking at anything that’s popular and posting some kind of troll comment, just to get a reaction, so they’ll just use a slur, or call someone a name, or post a shock image [...] you start to get a lot of really negative trolling”, reported one moderator [RM015], and others [RM005, RM008] also reported Front Page posts attracting increased trolling. Being nominated as subreddit of the day ‘advertises’ a subreddit to the wider community and pushes traffic in its direction, with the same consequences. Weninger et al (2013) have shown that postings which reach the Front Page, and which tend to have a large amount of comments already, receive even more comments as they become more visible. As moderators reported that when participation in a subreddit increased, poor quality comments were more of a problem than poor quality posts, the additional visibility of posts could be a real issue.

On the plus side, the fact that there are several types of event likely to trigger upsurges in traffic, some of which can be planned for in advance, means that there is a bank of reddit moderators with experience of ‘living through’ such surges who can not only jump in and help when they see them happening in other forums, but can also anticipate them and offer that help. An unexpected advantage of moderating r/ebola through the height of the outbreak was that it enabled one moderator, who had originally come across from r/books to offer experience of moderating a large forum (with, at the time, nearly 2.5 million subscribers, which had grown to 11 million by 1<sup>st</sup> January 2017), to build up extensive experience of the language, tactics and behaviour of racist posters, including usernames of known racists, which she was then able to offer to r/starwars when the movie subreddit began to receive racist posts following

the casting of a black actor in a leading role. The existing moderators of that subreddit, “didn’t even realise there were good options to control the racist language that people were using” [RM015]. Since then, “I’ve come into a couple of places, because I see the crisis happening and I come in and set automoderator, and set it up for them”. She has gained a reputation across the moderator metacommunity as someone who can help subreddits on any topic tackle abuse. And experienced moderators do look out for struggling subreddits: “When [a sudden increase in traffic] happened, the number of redditors stepping forward to help rose as well”, reported RM007, who had been on the receiving end of such help.

Moderators who hadn’t personally experienced a sudden increase in traffic generally speculated that if it did, they would deal with it by increasing the moderation team and this in turn would be done by drawing on their pool of existing users. In practice, this pool of users is deep, and often comes from across the reddit community rather than from inside the community under pressure. “[A] lot of [the moderators I approached] were people I was familiar with from somewhere else, or another moderator was familiar with ... we brought them in specifically because we needed people to deal with the moderation queue. Finding people who are interested in the topic when you’re scaling up, that’s really easy, because you [would] notice people who were consistently giving good information [...] but the technical side... it depends on who you know, and if you even know that’s an option to you” [RM015].

The interviews I conducted suggest that the larger a subreddit is, the more likely its moderators are to interact with the moderator community at large. As a subreddit grows, the volume of work required increases and moderators tend to take on specific roles within the group. Certain moderators develop and specialize in niche skills which will then become valuable to other subreddits. This dynamic may go unrecognized by a single moderator of a small subreddit

who has had no reason to expand or differentiate the moderator task, but the advantage of the reddit community is that the experienced moderators whose help may be needed are themselves able to spot a potential crisis emerging and offer help before their less experienced colleagues even realise it is needed. In practice, this is exactly what happened in the case of r/zika, with the r/ebola moderators setting it up to be ready to deal with an upsurge in interest even though this has not – so far – materialised. This is a valuable aspect of the reddit community and warrants further research in future.

Finally, we should not forget the lesson that moderating a subreddit during a PHEIC can be an emotionally challenging experience. At the height of the Ebola outbreak, many of the posts on r/ebola were coming from the ‘front line’, from healthcare workers in Ebola treatment centres, from victims themselves, their families and survivors. RM008 described this as “a soul scarring period of my life ... a time out was needed”. Reddit is not only able to scale up – it can also scale down, collectively and individually. The fact that the r/ebola subreddit was (and is) maintained by a moderation team allowed for any one of that team to take time out when needed, and this takes away any guilt the moderator may feel about removing a valuable resource at a time it is most needed. How platforms might provide better support in such circumstances should be considered; during a PHEIC professional counselling and psychological support may be required to deliver an appropriate duty of care to moderators at particularly stressful or harrowing times.

The egalitarian nature of moderation on reddit allows new moderators to come and go without significant changes to the subreddit’s character as there is less likely to be one dominant personality or ‘style’ than would be the case for an individual blog. In the case of 14 of the 18 reddit moderators I interviewed, their subreddit had existed before them will most likely continue after they no longer have time to maintain it. Subreddits are bigger than the sum of their

parts and this is an important lesson: not only do they create the spaces in which Collective Intelligence can emerge, but exactly what is *Created*, rather than *Collected* – the forum itself rather than its contents – comes extremely close to being the ‘clever mechanism’ Surowiecki suggested is needed to harness the Wisdom of the Crowd.

## 9. DISCUSSION AND CONCLUSIONS

---

### 9.1 Lessons identified from r/ebola

The Ebola virus PHEIC of 2014-2016, while tragic, provided a unique opportunity to research the information requirements of people experiencing a public health emergency and to explore how one specific information channel – the aggregator site reddit and its subreddit r/ebola – met those requirements. This elevated the study from its original aim in September 2013 of being a theoretical anticipation of what might happen under such conditions to a real world, real-time examination of what did. I have been able to test some of the assumptions reached throughout the project within its lifetime and examine how lessons identified by r/ebola were applied to a second PHEIC subreddit<sup>12</sup>, r/zika. The outbreak has also allowed me to test the quality of real information posted on a real platform – r/ebola – by asking doctors to rate it.

This has enabled me to make the following unique contributions to knowledge, which I will discuss further throughout this chapter:

[1] **Far at Risk, Near at Risk, Real at Risk**

I have identified three stages through which individuals experienced the outbreak, and in which those individuals exhibited distinct information requirements: Far at Risk, Near at Risk and Real at Risk.

[2] **The ideal Collective Intelligence Genome of a suitable platform is different in these different phases**

The different requirements in each stage are best served by different types of platform: in the early, Far at Risk stage by *Collection* of

---

<sup>12</sup> An outbreak of Polio in Syria in 2014 was also declared a PHEIC, but was not covered in this study as the ready availability of a vaccine to most of the world's population does not constitute the same level of potential public health emergency of emerging diseases such as Ebola and Zika.

information, but in the Near-at-Risk and Real-at-Risk this changes to *Collaboration*, where the community works together to answer questions, provide advice and support. *Collaboration by Voting*, to signpost the best material and ensure it is more visible and credible to other users, has value in all phases.

[3] **Credibility markers on reddit guide users to the better information**

Credibility signposts can and do highlight the information more likely to be of high quality. When doctors were asked to rate answers given on r/ebola, their ratings indicate that community voting scores do guide users away from bad information, and towards the better information (though this may not be the best information that could be obtained under ideal circumstances).

[4] **Quality of information is dependent on ‘Expert Moderators’**

The skills needed to moderate serious disease outbreak forums is less dependent on subject matter expertise (which can be provided by the user community itself) and more dependent on technical skills that can automate moderator functions. This can relieve pressure on a resource-stretched moderator team when traffic to the forum is increasing rapidly. Skills and skill-sets can be distributed across a moderator team, with some moderators taking on specialist roles.

[5] **Setting up a PHEIC forum from scratch may not be practical**

The complex characteristics that have been identified in r/ebola would be hard to implement from scratch. A more effective approach would be to identify an existing platform – such as reddit – with the most appropriate CIG for the context and use it to host the PHEIC discussions.

## 9.2 Limitations of the study

I recognise that this study and its conclusions is limited by small sample sizes and the fact that my interviewees came from a narrow demographic – NGO workers in Liberia and Sierra Leone, who may have held elite status that elevated them above the populations most at risk (Mitra and Van Delinder, 2007; King 2015) and the largely U.S.-based users of reddit, who were far away from the geographic centres of the outbreak. However, the quality of the interviewees and survey respondents was high: they were real people, experiencing a real PHEIC in real time. The information on r/ebola was assessed by doctors who, despite the pressures on their time, were prepared to assess the information according to their professional wisdom. A larger study could be conducted at a later stage, interviewing more people, but this would be subject to biases of reported behaviour (Alshenqueeti, 2014). I did not want the study opportunity offered by the Ebola PHEIC to pass.

It is likely that elite groups will be more likely than others to be reddit users already. However, as a PHEIC is more likely to emerge in an income-poor country context, a key policy consideration is how the knowledge gained by elite groups might be passed on to groups with less literacy and less digital literacy. Future work might explore mixed media approaches, e.g. combining a reddit user station with a public notice board or blackboard as used in the Ebola outbreak, or combining it with an interactive SMS or radio service.

I also acknowledge that my study does not attempt to suggest ways in which new users could be pushed towards reddit during a PHEIC, but instead assumes that reddit is already widely used and is likely to be acceptable to an at-risk population. In June 2017, Reddit was the 4<sup>th</sup> most popular website in the U.S. and the 6<sup>th</sup> most popular in the U.K. (Alexa.com, 2017), but I am still making assumptions, and I recognise this. More research will be necessary to

determine if the findings from these case studies are likely to be applicable across future PHEICs and public health emergencies in general.

### **9.3 The information journey and dimensions of risk**

The interviews I conducted with the NGO workers in West Africa suggest that the information journey taken by an individual in an affected population is to first become aware of the outbreak through general news channels they already use. This was upheld by the interviews with the moderators of the r/ebola and r/zika subreddits, who reported their interest, and initial posting activity, starting in r/worldnews, a subreddit focused on aggregating news stories through link posts to international news media. As interest grows, the information seeker often begins to search out information more actively, first looking for more scientific and fact-based information. The NGO workers looked to WHO and CDC, as well as for special broadcasts or programmes from trusted media such as the BBC. The reddit users looked for a dedicated subreddit – r/ebola – on which the same sources were strongly represented in the early posts: 16% of posts during r/ebola’s first active month, July 2014, came from intentional or national health agencies, with a further 43% from professional international media. A high interest in news reports from local African media on r/ebola, even though most users were U.S. based, points to an interest in local reporting on what is happening to communities closer at risk than one’s own. This would be an interesting subject for future study.

The type of information required, and how users wanted to absorb it was not static but changed throughout the period of the outbreak depending on the context. The NGO workers I interviewed in Sierra Leone and Liberia described a pattern of information seeking I labelled ‘Far at Risk’, ‘Near at Risk’ and ‘Real at Risk’. This was also observed on the r/ebola forum.

### ***Far at Risk***

‘Far at Risk’ people, who had no real chance of catching Ebola but were watching news items on the outbreak from cities and even countries where no cases had yet been recorded, wanted general information on how the disease spread and how many cases were being recorded.

In this regard, r/ebola was the ideal platform: an aggregator site that simply collected links to such information in one place for easy access, where reliable and non-sensationalised information on Ebola could be deposited and retrieved. There was little requirement in this stage for discussion or debate. This model does not remain optimal as the outbreak progresses, however.

### ***Near at Risk***

People living in regions where cases of Ebola were being reported needed to start being more careful about where they went and what they did. They began to exhibit a different style of information seeking I labelled ‘Near at Risk’, characterized by a voracious appetite for absorbing a greater volume of information from a wider range of sources: scientific; international, national and local media; social media; professional and informal. Some of these tended to be more speculative or sensationalist, requiring not only an aggregator site, but better ways to differentiate the ‘good’ information from the ‘bad’. On r/ebola, this was enabled by ‘flair’, which differentiated information from more reliable sources (e.g. CDC, WHO) from the less reliable ones (e.g. media, self). Strong upvoting from a well-informed community is also able to help. Though any voting system is in danger of being hijacked by conspiracy theorists and alarmists (or people who are just wrong), in the case of r/ebola this was tempered by well-informed users who, for example, countered claims that Ebola was likely to become airborne.

During this stage, those I interviewed wanted to correlate and compare information from multiple sources to form a version of events they felt they could trust. They readily acknowledged that certain sources may compromise accuracy for speed, but information was needed immediately: its veracity could be checked later. This demonstrated a utilization of the Wisdom of Crowds: if several sources were reporting the same thing, the receivers assumed the information was more likely to be correct than information they found in only a single report. Likewise, if each report contains part of the picture, consulting as many as possible may help a more complete picture to be constructed.

On r/ebola, the emergence of 'Near at Risk' behaviour coincided with the first cases of Ebola on U.S. soil and, in the West Africa Group, a key trigger event was Western NGO workers becoming infected. These people were in extremely high-risk categories (healthcare workers who had been in direct contact with Ebola victims in medical facilities, where some cases of infection may have been considered inevitable) as were the only victims of onward transmissions outside of Africa (two healthcare workers in the U.S. and one in Spain) but it was the reports of these infections that heightened concern amongst the West Africa Group and triggered the steepest increase in traffic, posting behaviour and subscriptions to the r/ebola forum. This suggests that individual and community perception of what is 'close to home' and puts one 'at risk' may be very different from the evaluation of epidemiologists. A clear explanation of why healthcare workers are at risk, and that such infections are not necessarily the result of a failure to manage and contain the disease competently should be a key message to have ready during future events.

The 'Near at Risk' stage saw information seekers wanting answers more specific to their situation, such the number and location of cases in their

locality. They sought clarification on terms or statements in the media and official sources that they did not fully understand. On r/ebola, this resulted in a sharp increase in the number of self-posts (from around 10% in July 2014, to nearly 40% in October) that required answers from other members of the community. Being aware of this likely shift and having subject matter experts on hand to answer questions when they come, will be of value in future PHEICS, as the style of question if not the exact components may well be echoed in future disease outbreaks. Future research might consider the role of text mining and automated answers to supplement the human responses.

An increase in ill-informed and speculative posts at this stage was a source of frustration to both the moderators and longer-term users of the r/ebola forum but seems symptomatic of the requirements of the ‘Near at Risk’ phase – which some of the new r/ebola users clearly considered the U.S. was moving into. People begin asking questions precisely because they are not already well-informed, but this needs careful monitoring to ensure information quality is kept high. This can place a huge time burden on the moderation team, stretching their resources and increasingly their reliance on the programming skills needed to make best use of any automated functions available.

In West Africa, radio phone-in talk shows allowed people to ask questions and have them answered. Discussion forums such as r/ebola are good alternatives for those who have access to them and future research could consider how they might be linked with radio phone-in shows or other media for the benefit of all. Reddit’s ability to run question and answer sessions, and to prepare FAQs for wider distribution, proved particularly valuable. A permanent FAQ, constantly updated on r/ebola<sup>13</sup>, was picked up by international newspapers.

---

<sup>13</sup> See [https://www.reddit.com/r/ebola/wiki/index#wiki\\_frequently\\_asked\\_questions](https://www.reddit.com/r/ebola/wiki/index#wiki_frequently_asked_questions) (accessed 7th July, 2015)

### ***Real at Risk***

‘Real at Risk’ – those in genuine danger of catching Ebola if they did not take deliberate action to avoid doing so, and/or who had close friends, colleagues or family who had contracted the disease – demonstrated different information requirements again. This shifted from collection of information to a requirement for a community space where information could be discussed and debated and where the information seeker wanted help in deciding what day-to-day actions to take. It is only in this phase that platform characteristics mirror that of a ‘typical’ health forum: a support community presenting mostly self-posts, where expert patients (Wilson, 2001) and the newly-diagnosed offer one another support, advice and guidance on living with their condition and circumstances.

The questions people ask in this latter phase – demonstrated by the requirement for information on the interaction of stray dogs with dead bodies on the streets of Liberia, and of concern over the safety of travelling on crowded buses, both of which were asked by the NGO workers and redditors – may not always be the ones public health officials are likely to anticipate, nor the ones distant policymakers are best placed to answer. An open discussion forum in which questions can be asked and answered directly by people with personal experience of the context is essential.

#### **9.4 Context Shift and Community Shift**

In these earlier phases – Far at Risk and Near at Risk – the aggregator characteristic of reddit, the purpose for which it was initially designed, served the information needs of the at-risk communities well. Information seekers entered the information chain through routine TV and radio news broadcasts, newspapers and subreddits such as r/news or r/science. They migrated to

dedicated sources of Ebola information – the websites of WHO and CDC, and the specific subreddits of r/ebola and r/zika – when they wanted more detailed information, generally correlating to when they felt the risk to them personally increased. In the case of r/ebola, this changing information requirement morphed the characteristics of the subreddit from a purely factual repository of scientific information, to a newsfeed with more in common with the r/worldnews subreddit from which it had emerged, with an increasing percentage of speculative media articles and a need for a more interactive forum on which users asked questions on specific cases and contexts. An important lesson to carry forward to future PHEICs is how r/ebola maintained credibility and trust in this context.

## **9.5 Credibility and trust on reddit**

At the beginning of my study period, I set out to explore the extent to which information could be, and should be, trusted by discussion forum readers. Reddit maps comfortably onto Wathen and Burkell's model of how users judge the credibility of online information. *Surface credibility* is provided by the professional design of the reddit platform, the organisation of information into the easily navigable layers of subreddits and the positioning of the subreddit to which the information was submitted (r/science, r/worldnews, r/ebola, r/conspiracy, etc). *Message credibility* is provided by the flair assigned to both the original source of the information and, if different, to the redditor who posted it and the karma score of that redditor. Voting on the messages shows how well others have evaluated the same information; information that has been evaluated well by the community at large is more likely to be trusted by individual community members. This enables reddit to 'fast track' its users through the model to the *content evaluation* stage, fulfilling its original intention to be the 'Front Page of the Internet' that guides its users to the best of the myriad content available (Singer et al, 2014).

The fact that on reddit, *content evaluation* is not a solitary exercise is an important characteristic of the platform. The collective evaluation by the reddit community is visible to all through the voting score, the ordering system and in the comments made on the original post. A reddit user who may not have the health literacy or experience to evaluate the content themselves does not have to: the community, which contains experts and experienced users, does it for them. Chapter 6 has shown that the credibility signposts do guide users to the better information available and away from bad information.

This adds to the current literature on the credibility of online information by providing a better understanding of not only how discussion forums use credibility signposts but also of how effective these credibility signposts are, which I verified through two surveys with three groups of doctors.

The use of effective and well-understood credibility differentiators within a website may also be a way to enable controversial information to coexist with more officially aligned information within a structured framework. Existing literature on health information suggests that ‘censoring’ controversial information and stifling debate can push some users towards online spaces supporting the more extreme positions, particularly if these are the only places where non-conventional views can be discussed (Grant et al, 2015). There would be great benefit in more research on the impact of allowing conspiracy theories and speculative media to remain visible but to be ‘flaired’ as such, showing that the community is sceptical of the information or rejects it. Research on how such information, and indeed feedback to the poster, is absorbed and processed would be useful.

The reddit community collectively evaluates the information submitted to the platform and *Decides* by *Voting* which is the best, demonstrating that Collective Intelligence can emerge from such forums. As reddit has grown and

diversified over time – from a portal that aggregates content hosted on external sites to a creator and host of original information (Leavitt, 2015; Kassing et al, 2015) – some individual subreddits have evolved different Collective Intelligence Genomes from reddit’s primary CIG. The complete site is a *Collection* – of items of information on which users collectively *Decide by Voting* which is of most value – but health subreddits are more often a *Collaboration*, between community members who share advice and experience and offer one another support. In the earlier stages of a PHEIC, a platform that can host information is sufficient; it is only in the later stages that the forum requires the same characteristics as a typical online health forum where users (can) *Collaborate* to discuss and debate the information and *Create* a more complete answer than has appeared in any single item they will find in the existing collection.

While some research has been carried out on how reddit’s character has changed, and how different types of posts have become more, or less prominent over time (Singer et al, 2014; Leavitt, 2015), more research is needed on how these evolving characteristics suit each subreddit’s intended purpose best. This will further help to suggest which characteristics are most appropriate during a public health emergency, including whether existing subreddits provide the most suitable channels, or whether – and how – a better one should be initiated.

There are disadvantages facing any PHEIC platform, however, which need to be fully acknowledged. The characteristics of health forums – on reddit and elsewhere – include a tendency towards expert patient moderators (Wilson, 2001; Dhatariya, 2006; Wadley et al, 2014), slow growth and strong emphasis on support and advice from a peer network of people who are living with, or have experience of dealing with, the conditions under discussion. Such people may not be readily available in a PHEIC, and while I have made some

suggestions of how a platform might deal with this, further research is needed on how this expertise can be provided if no-one has yet had the opportunity to develop it. Quality may inevitably drop at the time it is most needed and, if so, the ability of the community to sort and order information, and to allow the best to rise to the top when the percentage of 'good' information overall is lower, will be especially valuable.

It has been outside the scope of my study to see if the quantity of material evaluated as poor was larger when the volume of traffic was higher but this would be a useful area for further research.

The transition to a more self-post-heavy community is, for the most part, yet to happen on r/zika: most posts are still links to international health agencies including CDC and WHO, the U.S. National Institute of Health and university research pages. The characteristics of r/zika still very much match what is required and sought out in the early Far at Risk phase, with little discussion and very few self-posts; this may also be influenced by the fact that most of those who are genuinely Near- or Real-at-Risk are likely to be Spanish and Portuguese speakers who may be less likely to use reddit. In early 2017, Zika discussion across reddit was also taking place in r/science, r/worldnews, r/health, r/travel, r/politics r/brasil and others. If the interest of those forums' users increases, r/zika provides the additional information they are likely to require in the format they are likely to want it.

If traffic to r/zika grows, however, this may bring with it a broader range of information from a more diverse range of sources; if the moderators want to keep it as a knowledge subreddit, protected from the sensationalism and speculation they consider marred r/ebola, an additional subreddit more suited to the Near at Risk requirements – for example, r/zikanews – may be required.

By the time there is sufficient need for a platform to support a Real-at-Risk community, through which users can access people like themselves with real experience of living through past outbreaks, survivors or at least experienced frontline medical staff, may have emerged. They may be able to take on the role of subject matter expert, contributing valuable content to an appropriate subreddit while an experienced moderator guides the technical aspects, rules and norms of the forum. What is needed in the Real-at-Risk phase is not only access to information, but also the social capital that online networks can provide (Bourdieu, 1986; Kleine 2010; Hsieh et al, 2013).

There is extensive literature on how such communities form, how they are managed and how new members join, but less about how people take information from them and process it, or how they locate and utilize experts within new networks. This is another area where further research is needed.

A key conclusion that can be drawn from this study is the importance of different types of information, from different sources, at different points in the outbreak. This requires either different platform characteristics (different CIGs) at different stages, or a platform that is flexible enough to adapt from one to another as the outbreak progresses: pushing out dry facts, statistics and scientific description at the beginning of the outbreak but becoming more conducive to interaction and discussion, providing answers and offering support later (while still being able to push out dry facts). Reddit's flexibility to steer the r/ebola forum through such changes turned out to be largely dependent on the experience and skills of its moderator(s).

## **9.6 The Expert Moderator**

The final contribution to knowledge made by my study is the recognition of the role in quality assurance of the 'Expert Moderator', who needs a range of

moderation skills including technical and programming ability, experience of the platform's norms, conventions and history, and some subject matter expertise. These skills can be distributed between a team of moderators, some or all of whom take on specialist roles. The experience and technical skill of the moderator(s) is what most helps to keep the quality of information high in PHEIC subreddits, as knowing how to configure automoderation tools and how to ban posts containing certain words or from certain users, significantly eases the time burden imposed on moderators who need to weed out bad information when traffic is increasing rapidly. Context experience is also important: knowing what negative behaviour is likely to be encountered (and how to ban those exhibiting it quickly), who the site-wide trolls and troublemakers are and what challenges they are likely to bring. This experience can come from moderating any subreddit on any subject and does not tend to rely on specific subject-matter expertise: members of the moderator metacommunity hold transferable skills which can be as applicable to a subreddit on one topic as to one with a completely different focus. The moderators of subreddits on books, healthy eating, computer games and technology all brought invaluable skills with them to r/ebola while, as shown in Chapter 8, a subject matter expert with little moderating experience upset the group dynamics of the moderation team.

Subject matter experts are better utilized as content contributors, who can provide links to reputable and accurate information, answer questions posed in self-posts knowledgeably, correct inaccuracies, counter conspiracy theories and fill the information gap that can emerge when professional agencies are slow to react. The access to back-end technical functions afforded by formal moderator status is not necessary for this, and appointment to the formal role adds little to the value subject matter experts can provide. Misinformation and uncertainty is common when traffic is increasing most rapidly: this needs to be responded to with rational and measured comments. At such times, seeking

out subject matter experts who can become approved posters is valuable, but new moderators need technical skills and forum experience more. The subject matter expert's time is better left free for providing such expertise than becoming bogged down in routine maintenance and management of the forum.

This identification of the differentiation of roles is valuable not only for understanding how to configure the technical aspects of the system, but also its human elements – the moderation team (who may each bring different skills to the table) and the approved content contributors – during a future health emergency. It may be no coincidence that the size and demographics of reddit moderation teams I encountered map well against the characteristics of the Collectively Intelligent groups studied by Woolley et al (2015), including size (2-5 members), mixed gender and a relatively flat hierarchical structure.

## **9.7 Conclusions**

My study has offered several answers to my original research question of how and in what ways the technology affordances of online health forums benefit or hinder health-information seeking online during a public health emergency.

Firstly, it has shown that the quality of health information found in such forums can be high, though it is not perhaps always as high as it could be. Information that is rated highly by laypeople is given middling scores by doctors, though there is a strong correlation 0.67 ( $p < 10^{-5}$ ) overall between reddit and the average score of all doctors when information is assigned a rating of 'good', 'neutral' or 'bad'.

The ratings given by the reddit community are visible to other reddit users and thus health information seeking is helped by such signposts and markers, as well as others including the positioning of the subreddit, flair and the ability to

check the user home page of those posting the information. These do indeed guide users to the best of the information available. Furthermore, the technical tools supplied by reddit, which reduce the burden on the moderators' time, enhance their ability to remove poor quality information and enable them to ban those who consistently submit it. This points to such functions being extremely useful platform characteristics.

My research also answered a second question, which was not anticipated at the start of the study: at what stage of the outbreak is a discussion platform most appropriate? The answer is during the middle and later stages of the outbreak, when the concerns of those involved begin to turn inwards and focus on their own personal concerns rather than the dry facts, figures and statistics provided by scientists and public health agencies. During this later period, discussion forums not only provide answers to health information seekers' questions, but can also inform public health officials of what concerns the affected public has, which may not always be easy to anticipate, and enable discussion and debate around proposed interventions.

By focusing on the concept of Collective Intelligence, I have shown that the 'Collective Intelligence Genome' can be used to identify which, out of the many ways social media platforms can be configured, is likely to be the best suited to the information needs required during a PHEIC. As these needs can change as the context and affected communities change, a platform either needs to be able to adapt its CIG, or for there to be different platforms, each with its own CIG, to suit the evolving situation. Properly configured, such a platform can guide users through Wathen and Burkell's model to the information they are correct to trust.

The emergence of r/ebola as a source of reasonably high-quality, valued health information during the 2014-16 West Africa Ebola outbreak (for those who

have access to the Internet and sufficient literacy to use it) should not, in retrospect be a surprise. Reddit's dual nature as a news aggregator and a collection of communities allowed it to fit the changing needs of health information seekers as the outbreak evolved and the perceived risk from it changed. Studying what made it work, and what challenges it faced, provides a plethora of information on how a similar platform might be configured to deal with information needs during a future public health emergency.

This thesis has not, of course, been exhaustive, and there are several areas where further research is likely to prove valuable. There is much to be learned about how health information seekers absorb and process the information they find in online discussion forums, particularly around how susceptible they are to poor quality information that may slip through the net. Why are qualified healthcare professionals more likely to think laypeople will act inappropriately on poor quality information than laypeople themselves, who seem rightly confident that poor information will be recognized and called out, and which was on r/ebola? This would be particularly interesting to study in the context of how the many credibility signposts on reddit might influence that process.

There is scope for further analysis on the thousands of posts and comments that have been submitted to r/ebola and r/zika since their inceptions. They represent a permanent record of a unique period, and of the concerns and attitudes of the people living through it. This is a rich seam to mine, and there is far more to be learned from it. The same is true of the motivations, behaviour and psychology of the reddit PHEIC and chronic condition forum moderators. They put in remarkable amounts of time for little recognition and virtually no reward; while some research on this is emerging, more is needed. Finally, the research has shown that the best time to set up a PHEIC forum is not once the PHEIC has started, nor is it a good time to recruit the moderation team. Knowing what is already in place, and how to bring it into use when

needed, is much more likely to be effective. Reddit provided an excellent model for the recent PHEICs and offers an equally effective platform on which to prepare for future ones, particularly if ways of linking it with other media such as radio phone-in shows are considered, thus increasing access to groups who might otherwise be excluded. This should be an immediate focus of any future research on this topic.

## 10: REFERENCES

---

- Abramowitz, S. (2014). Ten things that anthropologists can do to fight the West African Ebola epidemic. *Science, Medicine and Anthropology-Ebola Series*, September, 26.
- Abras, C., Maloney-Krichmar, D., & Preece, J. (2004). User-centered design. *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications*, 37(4), 445-456.
- Agamben, G., (2005). *The State of Exception as a Paradigm of Government*. Chicago: University of Chicago Press
- Air, M., Roman, S. A., Yeo, H., Maser, C., Trapasso, T., Kinder, B., & Sosa, J. A. (2007). Outdated and incomplete: a review of thyroid cancer on the World Wide Web. *Thyroid*, 17(3), 259-265.
- Aitamurto, T. (2016, January). Collective Intelligence in Law Reforms: When the Logic of the Crowds and the Logic of Policymaking Collide. In *System Sciences (HICSS), 2016 49th Hawaii International Conference on* (pp. 2780-2789). IEEE.
- Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In *Action control* (pp. 11-39). Springer Berlin Heidelberg.
- Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human decision processes*, 50(2), 179-211.

Akerkar, S., & Bichile, L. (2004). Health information on the Internet: patient empowerment or patient deceit? *Indian journal of medical sciences*, 58(8), 321

Alexa (2017). Top sites in the United Kingdom. [www.Alexa.com](http://www.Alexa.com), <http://www.alexacom.com/topsites/countries/GB>; last accessed 12 June 2017

Alshenqeeti, H. (2014). Interviewing as a Data Collection Method: A Critical Review. *English Linguistics Research*, 3(1), 39.

Àlvaro, S. (2014). IEML: A Project for a New Humanism. An Interview with Pierre Lévy. *Centre de Cultura Contemporània Barcelona*, 26 March 2014 <http://lab.cccb.org/en/ieml-a-project-for-a-new-humanism-an-interview-with-pierre-levy/>; last accessed 12 June 2017

Ancker, J. S., Carpenter, K. M., Greene, P., Hoffman, R., Kukafka, R., Marlow, L. A., ... & Quillin, J. M. (2009). Peer-to-peer communication, cancer prevention, and the Internet. *Journal of health communication*, 14(S1), 38-46.

Ancker, J. S., & Kaufman, D. (2007). Rethinking health numeracy: A multidisciplinary literature review. *Journal of the American Medical Informatics Association*, 14(6), 713-721.

Arrington, M. (2006), 'Breaking News: Condé Nast/Wired Acquires Reddit', [Techcrunch.com](http://www.techcrunch.com), 31 October 2006 <http://www.wired.com/2006/10/wired-digital-acquires-reddit/>; last accessed 12 June 2017

Australian Emergency Management Institute (2013). Communicating with People with Disability: *National Guidelines for Emergency Managers*.

- Baccarne, B. (2012) 'The wisdom of the judges: een comparatief onderzoek tussen in-house en crowdsourced selectieprocedures bij idea competitions' Ghent University. [http://lib.ugent.be/fulltxt/RUG01/001/894/958/RUG01-001894958\\_2012\\_0001\\_AC.pdf](http://lib.ugent.be/fulltxt/RUG01/001/894/958/RUG01-001894958_2012_0001_AC.pdf); last accessed 12 June 2017
- Baird, D. (2015), Mail Online soars past 200 million monthly browsers as newspaper sites bounce back. *The Guardian*, 20 February 2015
- Bansil, P., Keenan, N. L., Zlot, A. I., & Gilliland, J. C. (2006). Health-related information on the Web: results from the HealthStyles Survey, 2002-2003. *Prev Chronic Dis*, 3(2), A36.
- Baron, N. S. (2010). *Always on: Language in an online and mobile world*. Oxford University Press.
- Barthel, M., Stocking, G., Holcomb, J., & Mitchell, A. (2016). Nearly eight-in-ten Reddit users get news on the site. *Pew Research Center* 25 (2016).
- Beach, L. R., & Mitchell, T. R. (1978). A contingency model for the selection of decision strategies. *Academy of management review*, 3(3), 439-449.
- Beck, A. M. (1973). *The ecology of stray dogs: a study of free-ranging urban animals*. Purdue University Press.
- Benton, A., Ungar, L., Hill, S., Hennessy, S., Mao, J., Chung, A., ... & Holmes, J. H. (2011). Identifying potential adverse effects using the web: A new approach to medical hypothesis generation. *Journal of biomedical informatics*, 44(6), 989-996.

Bernhardt, J. M., & Felter, E. M. (2004). Online pediatric information seeking among mothers of young children: results from a qualitative study using focus groups. *Journal of Medical Internet Research*, 6(1), e7.

Biermann, J. S., Golladay, G. J., Greenfield, M. L. V., & Baker, L. H. (1999). Evaluation of cancer information on the Internet. *Cancer*, 86(3), 381-390.

Biggs, J. B. (1987). *Student Approaches to Learning and Studying. Research Monograph*. Australian Council for Educational Research Ltd., Radford House, Frederick St., Hawthorn 3122, Australia.

Blackmore, K. L., & Nesbitt, K. V. (2008, January). Identifying risks for cross-disciplinary higher degree research students. In *Proceedings of the tenth conference on Australasian computing education-Volume 78* (pp. 43-52). Australian Computer Society, Inc.

Bonabeau, E. (2009). Decisions 2.0: The power of collective intelligence. *MIT Sloan management review*, 50(2), 45.

Bourdieu, P. (1986). The Forms of Capital. In handbook of theory and research for the sociology of education. Edited by: Richardson J.

Bower, J. L., & Christensen, C. M. (1995). Disruptive technologies: catching the wave. *Harvard Business Review*.

Broadbent, S., and Galotti, M., (2015), Collective Intelligence: How Does it Emerge?, Nesta, June 2015 <http://www.nesta.org.uk/publications/collective-intelligence-how-does-it-emerge>; last accessed 12 June 2017

- Bross, J., Richly, K., Kohnen, M., & Meinel, C. (2012). Identifying the top-dogs of the blogosphere. *Social network analysis and mining*, 2(1), 53-67.
- Brown, S. A., Massey, A. P., Montoya-Weiss, M. M., & Burkman, J. R. (2002). Do I really have to? User acceptance of mandated technology. *European journal of information systems*, 11(4), 283-295.
- Brownlie, J., Peckham, C., Waage, J., Woolhouse, M., Lyall, C., Meagher, L., ... & Nicoll, A. (2006). Foresight. Infectious diseases: Preparing for the future. Future threats. *London: Office of Science and Innovation*.
- Buntain, C., & Golbeck, J. (2014, April). Identifying social roles in reddit using network structure. In *Proceedings of the 23rd International Conference on World Wide Web* (pp. 615-620). ACM.
- Burkell, J. (2004). Health information seals of approval: what do they signify? *Information, Communication & Society*, 7(4), 491-509.
- Cabinet Office (2013). Guidance Pandemic Flu. First published 20 February 2013. <https://www.gov.uk/guidance/pandemic-flu>; last accessed 12 June 2017
- Cabinet Office (2015), National Risk Register of Civil Emergencies, 2015 Edition, Cabinet Office, London 2015
- Carr, N. (2011). *The shallows: What the Internet is doing to our brains*. WW Norton & Company.
- Casebourne, I., Davies, C., Fernandes, M., & Norman, N. (2012). Assessing the accuracy and quality of Wikipedia entries compared to popular online

encyclopaedias: A comparative preliminary study across disciplines in English, Spanish and Arabic. *Epic, Brighton, UK. Accedido o*, 9(10), 2012.

Centers for Disease Control and Prevention (2004). Fact Sheet: Basic Information About SARS. Centers for Disease Control and Prevention. January 13, 2004.

Centers for Disease Control and Prevention (2014). Questions and Answers: Estimating the Future Number of Cases in the Ebola Epidemic – Liberia and Sierra Leone, 2014-15

Centers for Disease Control and Prevention (2015) 2014 Ebola Outbreak in West Africa – Case Counts.

Chambers, R. (1981). Rapid rural appraisal: rationale and repertoire. *Public administration and development*, 1(2), 95-106.

Chambers, R. (1992). *Rural appraisal: rapid, relaxed and participatory*. Institute of Development Studies (UK).

Choi, D., Han, J., Chung, T., Ahn, Y. Y., Chun, B. G., & Kwon, T. T. (2015, November). Characterizing conversation patterns in Reddit: From the perspectives of content properties and user participation behaviors. In *Proceedings of the 2015 ACM on Conference on Online Social Networks* (pp. 233-243). ACM.

Christensen, W., & Suess, R. (1978). Hobbyist computerized bulletin board. *Byte Magazine*, 3(11), 150-158.

Clark, A. & Chalmers, D. (1998). The extended mind. *Analysis*, 7-19.

- Clark, A. (2002). Towards a science of the bio-technological mind. *International Journal of Cognition and Technology*, 1(1), 21-33.
- Cohen, P. (2010). Scholars test web alternative to peer review. *The New York Times*, August 23, 2010.
- Cole, J. (2010). Interoperability in a Crisis. *Human Factors and Organisational Processes*. Royal United Services Institute, 2010
- Cole, J. (2013). Antimicrobial resistance, infection control and planning for pandemics: the importance of knowledge transfer in healthcare resilience and emergency planning. *Journal of business continuity & emergency planning*, 6(2), 122-135.
- Cole, J. (Ed) (2015). Serious Infectious Disease: Challenges for Security and Defence *RUSI/STFC Conference Series No.6* June 2015
- Cole, J. (2016). Antimicrobial resistance—a ‘rising tide’ of national (and international) risk. *Journal of Hospital Infection*, 92(3), e4.
- Cole, J. & Hawker, E. (2014). *Emergency Services Communications: Resilience for the 21<sup>st</sup> Century*. Royal United Services Institute, 2014
- Cole, J. & Watkins, C (2013). Discussion Group 1: Changing Behaviour in Antibiotic Prescribing in J. Cole (Ed) (2013). Tackling Antimicrobial Resistance: Identifying Future Research Themes *RUSI/STFC Conference Series No.1* August 2013
- Cole, J., & Watkins, C. (2015). International employees’ concerns during serious disease outbreaks and the potential impact on business continuity:

Lessons identified from the 2014–15 West African Ebola outbreak. *Journal of business continuity & emergency planning*, 9(2), 149-162.

Cole, J., Watkins, C., & Kleine, D. (2016). Health advice from Internet discussion forums: how bad is dangerous?. *Journal of Medical Internet Research*, 18(1).

Conole, G., & Dyke, M. (2004). What are the affordances of information and communication technologies? *Alt-J: Research in Learning Technology*, 12(2).

Coulson, N. S., & Shaw, R. L. (2013). Nurturing health-related online support groups: Exploring the experiences of patient moderators. *Computers in Human Behavior*, 29(4), 1695-1701.

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Creswell, J. W. (2014). *A concise introduction to mixed methods research*. Sage Publications.

Crocco, A. G., Villasis-Keever, M., & Jadad, A. R. (2002). Analysis of cases of harm associated with use of health information on the Internet. *Jama*, 287(21), 2869-2871.

Crystal, D. (2001). *Language and the Internet*. Cambridge University Press.

Dahlen, H. G., & Homer, C. S. (2013). ‘Motherbirth or childbirth’? A prospective analysis of vaginal birth after caesarean blogs. *Midwifery*, 29(2), 167-173.

Dalkey, N. (1969). An experimental study of group opinion: the Delphi method. *Futures*, 1(5), 408-426.

Das, S., & Lavoie, A. (2014, May). The effects of feedback on human behavior in social media: An inverse reinforcement learning model. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems* (pp. 653-660). International Foundation for Autonomous Agents and Multiagent Systems.

Davis Jr, F. D. (1986). *A technology acceptance model for empirically testing new end-user information systems: Theory and results* (Doctoral dissertation, Massachusetts Institute of Technology).

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 319-340.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: a comparison of two theoretical models. *Management science*, 35(8), 982-1003.

Dawood, F. S., Iuliano, A. D., Reed, C., Meltzer, M. I., Shay, D. K., Cheng, P. Y., ... & Feikin, D. R. (2012). Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: a modelling study. *The Lancet infectious diseases*, 12(9), 687-695.

Dearden, A., & Rizvi, H. (2008, October). Adapting participatory and agile software methods to participatory rural development. In *Proceedings of the Tenth Anniversary Conference on Participatory Design 2008* (pp. 221-225). Indiana University.

- DeLone, W. H., & McLean, E. R. (1992). Information systems success: The quest for the dependent variable. *Information systems research*, 3(1), 60-95.
- DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: a ten-year update. *Journal of management information systems*, 19(4), 9-30.
- Department of Education. (2015). Schools, Pupils and Their Characteristics. *Statistical First Release*, 11 June 2015
- Department of Health (2012), UK Pandemic Influenza Communications Strategy 2012, *Department of Health Gateway Reference 18415*, December 2012
- Dhatariya, K. K. (2006). Policing of information from Internet breast cancer list: Findings may not be generally applicable. *BMJ: British Medical Journal*, 332(7549), 1095.
- Diabetes UK (2015). State of the Nation Challenges for 2015 and Beyond. *London: Diabetes UK*. (WebCite<sup>®</sup>: <http://www.webcitation.org/6as2DvLUe>)
- Digg. (2015). Digg By the Numbers 2015 Edition. *Digg* 30 January 2015
- DiMaggio, P., & Hargittai, E. (2001). From the 'digital divide' to 'digital inequality': Studying Internet use as penetration increases. *Princeton: Center for Arts and Cultural Policy Studies, Woodrow Wilson School, Princeton University*, 4(1), 4-2.

- DiMaggio, P., Hargittai, E., Celeste, C., & Shafer, S. (2004). From unequal access to differentiated use: A literature review and agenda for research on digital inequality. *Social inequality*, 355-400.
- Diviani, N., Putte, B., Giani, S., & Weert, J. C. M. (2015). Low health literacy and evaluation of online health information: a systematic review of the literature. *Journal of Medical Internet Research*, 17(5).
- Dobransky, K., & Hargittai, E. (2006). The disability divide in Internet access and use. *Information, Communication & Society*, 9(3), 313-334.
- Donaldson, L. J., Rutter, P. D., Ellis, B. M., Greaves, F. E., Mytton, O. T., Pebody, R. G., & Yardley, I. E. (2009). Mortality from pandemic A/H1N1 2009 influenza in England: public health surveillance study. *Bmj*, 339, b5213.
- Duggan, M., (2014). Online Harassment. *Pew Research Center*.
- Duhaime, E. P., Olson, G. M., & Malone, T. W. (2014). Broad Participation in Collective Problem Solving Can Influence Participants and Lead to Better Solutions: Evidence from the MIT Climate CoLab. *IT Center for Collective Intelligence, Cambridge, MA*. [http://cci.mit.edu/working\\_papers\\_2012\\_2013/duhaime%20colab%20wp](http://cci.mit.edu/working_papers_2012_2013/duhaime%20colab%20wp), 206-2015; last accessed 12 June 2017
- Dunn, R., Giannitti, M. C., Murray, J. B., Rossi, I., Geisert, G., & Quinn, P. (1990). Grouping students for instruction: Effects of learning style on achievement and attitudes. *The Journal of social psychology*, 130(4), 485-494.
- Esquivel, A., Meric-Bernstam, F., & Bernstam, E. V. (2006). Accuracy and self-correction of information received from an Internet breast cancer list: content analysis. *Bmj*, 332(7547), 939-942.

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics*, 5(1), 1-4.

Eysenbach, G., & Köhler, C. (2002). How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. *Bmj*, 324(7337), 573-577.

Eysenbach, G., Powell, J., Kuss, O., & Sa, E. R. (2002). Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. *Jama*, 287(20), 2691-2700.

Eysenbach, G., Powell, J., Englesakis, M., Rizo, C., & Stern, A. (2004). Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. *Bmj*, 328(7449), 1166.

Feenberg, A. L., Licht, J. M., Kane, K. P., Moran, K., & Smith, R. A. (1996). The online patient meeting. *Journal of the neurological sciences*, 139, 129-131.

Fisher, K. E., & McKechnie, L. (2005). *Theories of information behavior*. Information Today, Inc.

Fishbein, M. (1979). A theory of reasoned action: some applications and implications. *Nebraska Symposium on Motivation*, 27, 65-116.

Fitzpatrick, K. (2011). *Planned obsolescence: Publishing, technology, and the future of the academy*. NYU Press.

Fleming, N. D. (2001). *Teaching and learning styles: VARK strategies*. IGI Global.

Fuchs, C., & Horak, E. (2008). Africa and the digital divide. *Telematics and informatics*, 25(2), 99-116.

Fung, W. K., & Philip, L. H. (2003). SARS case-fatality rates. *Canadian Medical Association Journal*, 169(4), 277-278.

Funk, S., Gilad, E., Watkins, C., & Jansen, V. A. (2009). The spread of awareness and its impact on epidemic outbreaks. *Proceedings of the National Academy of Sciences*, 106(16), 6872-6877.

Funk, S., Gilad, E., & Jansen, V. A. (2010). Endemic disease, awareness, and local behavioural response. *Journal of theoretical biology*, 264(2), 501-509.

Funk, S., & Jansen, V. A. (2013). The talk of the town: modelling the spread of information and changes in behaviour. In *Modeling the interplay between human behavior and the spread of infectious diseases* (pp. 93-102). Springer New York.

Galton, F. (1907). Vox populi (The wisdom of crowds). *Nature*, 75(7), 450-451.

Gao, M. (2016, March). Intelligent Interface for Organizing Online Social Opinions on Reddit. In *Companion Publication of the 21st International Conference on Intelligent User Interfaces* (pp. 134-137). ACM.

Garske, T., Legrand, J., Donnelly, C. A., Ward, H., Cauchemez, S., Fraser, C., ... & Ghani, A. C. (2009). Assessing the severity of the novel influenza A/H1N1 pandemic. *BMJ*, 339, b2840.

Gaver, W. W. (1991, April). Technology affordances. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 79-84). ACM.

Gaver, W. W. (1992, December). The affordances of media spaces for collaboration. In *Proceedings of the 1992 ACM conference on Computer-supported cooperative work* (pp. 17-24). ACM.

Ghamari-Tabrizi, S. (2000). Simulating the unthinkable: Gaming future war in the 1950s and 1960s. *Social Studies of Science*, 30(2), 163-223.

Gibson, J. J. (1979). The theory of affordances *The Ecological Approach to Visual Perception* (pp. 127-143). Boston: Houghton Mifflin.

Gibson, J. J. (2014). *The ecological approach to visual perception: classic edition*. Psychology Press.

Gilbert, E. (2013, February). Widespread underprovision on Reddit. In *Proceedings of the 2013 conference on Computer supported cooperative work* (pp. 803-808). ACM.

Giles, J., (2005), Internet encyclopaedias go head to head, *Nature* Vol 438:15, pp900-901

Glezen, W. Paul. Emerging infections: pandemic influenza. *Epidemiologic reviews* 18.1 (1996): 64-76.

- Golder, S. A., & Donath, J. (2004). Social roles in electronic communities. *Internet Research, 5*, 19-22.
- Goodman, A. B., Meites, E., Anstey, E. H., Fullerton, K. E., Jayatilleke, A., Ruben, W., ... & Kirkcaldy, R. D. (2015). Clinical inquiries received by CDC regarding suspected Ebola virus disease in children—United States, July 9, 2014–January 4, 2015. *Morbidity & Mortality Weekly Report, 64*, 1006-1010.
- Goody, J. (1977). *The domestication of the savage mind*. Cambridge University Press.
- Gordon, G. (1979, December). Finding out about child (0-5 years) feeding practices'. In *Conference Rapid Rural Appraisal, Institute of Development Studies, Sussex, UK (4-7 December, 1979)*.
- Grant, L., Hausman, B. L., Cashion, M., Lucchesi, N., Patel, K., & Roberts, J. (2015). Vaccination persuasion online: a qualitative study of two provaccine and two vaccine-skeptical websites. *Journal of Medical Internet Research, 17*(5), e133.
- Graybill, J. K., Dooling, S., Shandas, V., Withey, J., Greve, A., & Simon, G. L. (2006). A rough guide to interdisciplinarity: Graduate student perspectives. *BioScience, 56*(9), 757-763.
- Grindon, G., & Flood, C. (2014). *Disobedient objects*. V&A Publications.
- Groselj, D. (2014). A webometric analysis of online health information: sponsorship, platform type and link structures. *Online Information Review, 38*(2), 209-231.

- Green, J., & Jenkins, H. (2009). The moral economy of Web 2.0: audience research and convergence culture. *Media industries: History, theory, and method*, 213-225.
- Gupta, U. G., & Clarke, R. E. (1996). Theory and applications of the Delphi technique: A bibliography (1975–1994). *Technological forecasting and social change*, 53(2), 185-211.
- Halverson, C. A. (2002). Activity theory and distributed cognition: Or what does CSCW need to DO with theories?. *Computer Supported Cooperative Work (CSCW)*, 11(1-2), 243-267.
- Haselton, M. G., Nettle, D., & Murray, D. R. (2005). The evolution of cognitive bias. *The handbook of evolutionary psychology*.
- Hesse, B. W., Nelson, D. E., Kreps, G. L., Croyle, R. T., Arora, N. K., Rimer, B. K., & Viswanath, K. (2005). Trust and sources of health information: the impact of the Internet and its implications for health care providers: findings from the first Health Information National Trends Survey. *Archives of internal medicine*, 165(22), 2618-2624.
- Heywood, J., (1546) *A dialogue conteinyng the nomber in effect of all the prouerbes in the Englishe tongue*. Thomas Berthelet: London
- Hill, C. A., & O'Hara, E. A. (2006). A cognitive theory of trust. *Wash. UL Rev.*, 84, 1717.
- Hiltz, S. R., & Turoff, M. (1978). *The network nation: Human communication via computer*. Mit Press.

Hoch, D. B., Norris, D., Lester, J. E., & Marcus, A. D. (1999). Information exchange in an epilepsy forum on the World Wide Web. *Seizure*, 8(1), 30-34.

Hoffman, D. L., & Novak, T. P. (1998). Bridging the racial divide on the Internet. *Science*, 280(5362), 390-391.

Howard, P. N., Busch, L., & Sheets, P. (2010). Comparing Digital Divides: Internet Access and Social Inequality in Canada and the United States. *Canadian Journal of Communication*, 35(1), 109.

Hsieh, G., Hou, Y., Chen, I., & Truong, K. N. (2013, February). Welcome!: social and psychological predictors of volunteer socializers in online communities. In *Proceedings of the 2013 conference on Computer supported cooperative work* (pp. 827-838). ACM.

Huh, J., Yetisgen-Yildiz, M., & Pratt, W. (2013). Text classification for assisting moderators in online health communities. *Journal of biomedical informatics*, 46(6), 998-1005.

Huh, J. (2015). Clinical questions in online health communities: the case of see your doctor threads. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 1488-1499). ACM.

Hutchins, E. (1995). *Cognition in the Wild*. MIT press.

Hutchins, E. (2006). The distributed cognition perspective on human interaction. *Roots of human sociality: Culture, cognition and interaction*, 375-398.

Hwang, Y., Al-Arabi, M., & Shin, D. H. (2016). Understanding technology acceptance in a mandatory environment: A literature review. *Information Development*, 32(4), 1266-1283.

Hyer, R. N., Covello, V. T., & World Health Organization. (2005). *Effective media communication during public health emergencies: a WHO handbook*.

Impicciatore, P., Pandolfini, C., Casella, N., & Bonati, M. (1997). Reliability of health information for the public on the World Wide Web: systematic survey of advice on managing fever in children at home. *Bmj*, 314(7098), 1875.

International Telecommunications Union (ITU) World Telecommunication/ICT Indicators database 2015, 18th edition. Available at <http://www.itu.int/ITU-D/ict/publications/world/world.html>; last accessed 12 June 2017

Internet World Statistics (2014), Usage and Population Statistics, Africa, 2014, available at [www.internetworldstats.com/stats1.htm](http://www.internetworldstats.com/stats1.htm), last accessed 12 June 2017

Introne, J., Laubacher, R., Olson, G., & Malone, T. (2013). Solving wicked social problems with socio-computational systems. *KI-Künstliche Intelligenz*, 27(1), 45-52.

Janis, I. L. (1972). *Victims of groupthink: a psychological study of foreign-policy decisions and fiascoes*. Oxford, England. 277

JafariNaimi, N., & Meyers, E. M. (2015, February). Collective intelligence or group think?: Engaging participation patterns in world without oil.

In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 1872-1881). ACM.

Jenkins, H. (2002). Interactive audiences? The collective intelligence of media fans. *The new media book*, 157-76.

Jiang, A. H., Bischof, Z. S., & Bustamante, F. E. (2014, August). A cliq of content curators. In *ACM SIGCOMM Computer Communication Review* (Vol. 44, No. 4, pp. 139-140). ACM.

Kaptelinin, V., & Nardi, B. (2012, May). Affordances in HCI: toward a mediated action perspective. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 967-976). ACM.

Kassing, S., Oosterman, J., Bozzon, A., & Houben, G. J. (2015, April). Locating domain-specific contents and experts on social bookmarking communities. In *Proceedings of the 30th Annual ACM Symposium on Applied Computing* (pp. 747-752). ACM.

Kersey, B. (2012). Watch This Space: The Story of Reddit. *The Verge*, 30 November 2012

Kim, Y. (2016). Trust in health information websites: A systematic literature review on the antecedents of trust. *Health informatics journal*, 22(2), 355-369.

King, S. (2015). Increasing the power of the poor? NGO-led social accountability initiatives and political capabilities in rural Uganda. *The European Journal of Development Research*, 27(5), 887-902.

Kleine, D. (2010). ICT4WHAT? – Using the choice framework to operationalise the capability approach to development. *Journal of International Development*, 22(5), 674-692.

Koch-Weser, S., Bradshaw, Y. S., Gualtieri, L., & Gallagher, S. S. (2010). The Internet as a health information source: findings from the 2007 Health Information National Trends Survey and implications for health communication. *Journal of health communication*, 15(sup3), 279-293.

Lampe, C., & Resnick, P. (2004, April). Slash (dot) and burn: distributed moderation in a large online conversation space. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 543-550). ACM.

Lampe, C., & Johnston, E. (2005, November). Follow the (slash) dot: effects of feedback on new members in an online community. In *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work* (pp. 11-20). ACM.

Lanier, J. (2010). *You are not a gadget: A manifesto*. Vintage.

Latour, B. (1986). Visualization and cognition. *Knowledge and society*, 6(6), 1-40.

Latour, B. (1991). Technology is society made durable in *Sociology of monsters: Essays on power, technology, and domination*, John Law (ed), 103-131.

Lawrentschuk, N., Sasges, D., Tasevski, R., Abouassaly, R., Scott, A. M., & Davis, I. D. (2012). Oncology health information quality on the Internet: a multilingual evaluation. *Annals of surgical oncology*, *19*(3), 706-713.

Leach, M. (2015). The Ebola Crisis and Post-2015 Development. *Journal of International Development*, *27*(6), 816-834.

Leavitt, A. (2015, February). This is a throwaway account: Temporary technical identities and perceptions of anonymity in a massive online community. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 317-327). ACM.

Lévy, P. (1995). Pour l'intelligence collective. *Le monde diplomatique*, *25*.

Lévy, P. (1997). *L'intelligence collective: pour une anthropologie du cyberspace*. Paris: La découverte.

Lévy, P. (1997). *Collective intelligence*. New York: Plenum/Harper Collins.

Lévy, P. (2010). From social computing to reflexive collective intelligence: The IEML research program. *Information Sciences*, *180*(1), 71-94.

Linstone, H. A., & Turoff, M. (Eds.). (1975). *The Delphi method: Techniques and applications* (Vol. 29). Reading, MA: Addison-Wesley.

Linstone, H. A., & Turoff, M. (2011). Delphi: A brief look backward and forward. *Technological Forecasting and Social Change*, *78*(9), 1712-1719.

Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences*, 108(22), 9020-9025.

Lovink, G., & Rossiter, N. (2009). The digital given: 10 Web 2.0 theses. *Fibreculture Journal*, 14.

Luo, W., & Najdawi, M. (2004). Trust-building measures: a review of consumer health portals. *Communications of the ACM*, 47(1), 108-113.

Ma, L., Lee, C. S., & Goh, D. H. L. (2012, April). Sharing in social news websites: Examining the influence of news attributes and news sharers. In *Information Technology: New Generations (ITNG), 2012 Ninth International Conference on* (pp. 726-731). IEEE.

MacDonald, A. (1998). *Uncertain utopia: Science fiction media fandom and computer mediated communication* (Vol. 139). New Jersey: Hampton Press Inc.

Malone, T. W. (2006). *What Is Collective Intelligence and What Will We Do About it?* Edited transcript of remarks made at the official launch of the MIT Center for Collective Intelligence, October 13, 2006

Malone, T. W. (2008). What is Collective Intelligence and What Will We Do About it? *Collective Intelligence: Creating a Prosperous World at Peace, Earth Intelligence Network, Oakton, Virginia*, 1-4.

Malone, T. W. (2011). Climate CoLab, *Collective intelligence to address climate change*. Finalists chosen in world-wide contest.

Malone, T. W., & Bernstein, M. S. (2015). *Handbook of collective intelligence*. MIT Press.

Malone, T. W., & Crowston, K. (1990, September). What is coordination theory and how can it help design cooperative work systems? In *Proceedings of the 1990 ACM conference on Computer-supported cooperative work* (pp. 357-370). ACM.

Malone, T. W., Laubacher, R., & Dellarocas, C. (2009). Harnessing crowds: Mapping the genome of collective intelligence. *MIT Sloan Research Paper No. 4732-09*

Maloney, M. T., & Mulherin, J. H. (2003). The complexity of price discovery in an efficient market: the stock market reaction to the Challenger crash. *Journal of corporate finance*, 9(4), 453-479.

Maloney-Krichmar, D., & Preece, J. (2005). A multilevel analysis of sociability, usability, and community dynamics in an online health community. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 12(2), 201-232.

Marco, J., Barba, R., Losa, J. E., de la Serna, C. M., Sainz, M., Lantigua, I. F., & de la Serna, J. L. (2006). Advice from a medical expert through the Internet on queries about AIDS and hepatitis: analysis of a pilot experiment. *PLoS Med*, 3(7), e256.

McGrenere, J., & Ho, W. (2000, May). Affordances: Clarifying and evolving a concept. In *Graphics interface* (Vol. 2000, pp. 179-186).

- Merritt, E. (2012). *An analysis of the discourse of Internet trolling: A case study of Reddit. com* (Doctoral dissertation).
- Miles, J., Petrie, C., & Steel, M. (2000). Slimming on the Internet. *Journal of the Royal Society of Medicine*, 93(5), 254-257.
- Mitra, A., & Van Delinder, J. (2007). Elite women's roles in empowering oppressed women: Volunteer work in NGOs in Kolkata, India. *Recent studies on Indian women*, 355-380.
- Morreale, D., & Bertone, G. (2015). Alternate Reality Game and Cultural Heritage: The "Aequilibrium" Project. *Digital Studies/Le champ numérique*.
- Morse, F., (2015). Who Uses Reddit in the UK? *BBC Newsbeat* 24 June 2015 <http://www.bbc.co.uk/newsbeat/article/33254299/who-uses-reddit-in-the-uk>; last accessed 12 June 2017
- Munson, S. A., Cavusoglu, H., Frisch, L., & Fels, S. (2013). Sociotechnical challenges and progress in using social media for health. *Journal of medical Internet research*, 15(10), e226.
- Mursch, K., & Behnke-Mursch, J. (2003). Internet-based interaction among brain tumour patients. Analysis of a medical mailing list. *Zentralblatt für Neurochirurgie*, 64(02), 71-75.
- Nielsen, J. (2006). Participation inequality: lurkers vs. contributors in Internet communities. *Jakob Nielsen's Alertbox*.
- NHS England. (2013). Improving general practice – a call to action. *Evidence pack*.

- Nölke, L., Mensing, M., Krämer, A., & Hornberg, C. (2015). Sociodemographic and health-(care-) related characteristics of online health information seekers: a cross-sectional German study. *BMC public health, 15*(1), 31.
- Nonnecke, B., Andrews, D., & Preece, J. (2006). Non-public and public online community participation: Needs, attitudes and behavior. *Electronic Commerce Research, 6*(1), 7-20.
- Norman, D. A. (1986). Cognitive engineering. *User centered system design: New perspectives on human-computer interaction, 3161*.
- Norman, D. A. (1988). *The psychology of everyday things*. New York: Basic books.
- Norman, D. A. (1999). Affordance, conventions, and design. *Interactions, 6* (3), 38-43.
- Norman, D. (2002). Emotion & design: attractive things work better. *interactions, 9*(4), 36-42.
- Nguyen-Van-Tam, J. S., & Sellwood, C. (2013a). Intervention strategies for emerging respiratory virus infections: policy and public health considerations. *Current opinion in virology, 3*(2), 192-198.
- Nguyen-Van-Tam, J. S., & Sellwood, C. (2013b). Preparing for a potential A (H7N9) pandemic: lessons from the deployment of A (H1N1) pandemic vaccines. *Expert review of vaccines, 12*(8), 825-828.
- Ofcom (2014). *Adults' media use and attitudes Report 2014*. April 2014

Ofcom (2016). *Adults' media use and attitudes Report 2016*. April 2016

Omicini, A., Ricci, A., & Viroli, M. (2008). Artifacts in the A&A meta-model for multi-agent systems. *Autonomous agents and multi-agent systems*, 17(3), 432-456.

ONS Statistical Bulletin (2015). *Internet Access – Households and Individuals 2013*. The Office for National Statistics. 2015-08-18.  
(WebCite®: <http://www.webcitation.org/6as1KJvqP>)

ONS Statistical Bulletin (2016) *Internet Users in the UK*. The Office for National Statistics. 20 May 2016.

Pandey, A., Atkins, K. E., Medlock, J., Wenzel, N., Townsend, J. P., Childs, J. E., ... & Galvani, A. P. (2014). Strategies for containing Ebola in west Africa. *Science*, 346(6212), 991-995.

Park, D., Sachar, S., Diakopoulos, N., & Elmqvist, N. (2016, May). Supporting comment moderators in identifying high quality online news comments. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 1114-1125). ACM.

PharmaForward LLC (2013). *Most Searched for Diseases on Google*. 2013 May. (WebCite®: <http://www.webcitation.org/6as2trxdR>)

Preece, J., Abras, C., & Maloney-Krichmar, D. (2004). Designing and evaluating online communities: research speaks to emerging practice. *International Journal of Web Based Communities*, 1(1), 2-18.

Public Health England. (2014a) *Communicable Disease Outbreak Management: Operational Guidance*, prepared by McAuslane and Morgan, CIDSC, Public Health England, PHE publications gateway number 2014252

Public Health England. (2014b). *HIV in the United Kingdom: 2014 Report*.  
Public Health England. (WebCite®: <http://www.webcitation.org/6as2XGatN>)

Public Health Wales. (2015). *Chickenpox*. Public Health Wales. 2015-08-18.  
(WebCite®: <http://www.webcitation.org/6as2iXTfq>)

Pushdoctor (2015). *Last night the Internet saved my life Digital Health Report 2015*. Pushdoctor. (WebCite®: <http://www.webcitation.org/6as0FAW8K>)

Reddit (2016). *Results of the Reddit Demographics Survey 2016*.  
[https://www.reddit.com/r/SampleSize/comments/559ge8/academicthe\\_reddit\\_demographics\\_survey\\_2016\\_all/](https://www.reddit.com/r/SampleSize/comments/559ge8/academicthe_reddit_demographics_survey_2016_all/) and <http://imgur.com/gallery/cPzIB>

Reynolds, B., & Quinn, S. C. (2008). Effective communication during an influenza pandemic: the value of using a crisis and emergency risk communication framework. *Health Promotion Practice*, 9(4\_suppl), 13S-17S.

Reynolds, B., & W. Seeger, M. (2005). Crisis and emergency risk communication as an integrative model. *Journal of health communication*, 10(1), 43-55.

Rivera, I., Warren, J., & Curran, J. (2016, February). Quantifying mood, content and dynamics of health forums. In *Proceedings of the Australasian Computer Science Week Multiconference* (p. 67). ACM.

Roberts, J. (2015). Mumsnet Founder Reflects on Growth of ‘Fifth Child’, *the Guardian Online*, 21 March 2015

Robinson, P. (1999). The CNN effect: can the news media drive foreign policy?. *Review of international studies*, 25(02), 301-309.

Robinson, P. (2000). The policy-media interaction model: Measuring media power during humanitarian crisis. *Journal of Peace Research*, 37(5), 613-633.

Rowe, G., Wright, G., & Bolger, F. (1991). Delphi: a reevaluation of research and theory. *Technological Forecasting and Social Change*, 39(3), 235-251.

Salomon, G. (1993). No distribution without individuals’ cognition: A dynamic interactional view. *Distributed cognitions: Psychological and educational considerations*, 111-138.

San Norberto, E. M., Taylor, J., Salvador, R., Revilla, Á., Merino, B., & Vaquero, C. (2011). The quality of information available on the Internet about aortic aneurysm and its endovascular treatment. *Revista Española de Cardiología (English Edition)*, 64(10), 869-875.

Sanger, L. (2005). The early history of Nupedia and Wikipedia: a memoir in *Open sources, 2.0: The continuing evolution*. O’Reilly Media, Inc. 307-38

Schmidt, K., & Bannon, L. (1992). Taking CSCW seriously. *Computer Supported Cooperative Work (CSCW)*, 1(1-2), 7-40.

Schulz, P. J., & Nakamoto, K. (2011, March). “Bad” Literacy, the Internet, and the Limits of Patient Empowerment. In *AAAI Spring Symposium: AI and Health Communication*.

Schulz, P. J., & Nakamoto, K. (2013). Health literacy and patient empowerment in health communication: the importance of separating conjoined twins. *Patient education and counseling*, 90(1), 4-11.

Schuurman, D., Baccarne, B., De Marez, L., & Mechant, P. (2012). Smart ideas for smart cities: Investigating crowdsourcing for generating and selecting ideas for ICT innovation in a city context. *Journal of theoretical and applied electronic commerce research*, 7(3), 49-62.

Schwartz, K. L., Roe, T., Northrup, J., Meza, J., Seifeldin, R., & Neale, A. V. (2006). Family medicine patients' use of the Internet for health information: a MetroNet study. *The Journal of the American Board of Family Medicine*, 19(1), 39-45.

Seifert, C. M., & Hutchins, E. L. (1992). Error as opportunity: Learning in a cooperative task. *Human-Computer Interaction*, 7(4), 409-435.

Sen, A. (1997). Maximization and the Act of Choice. *Econometrica: Journal of the Econometric Society*, 745-779.

Sen, A. (1999). *Freedom as development*. Oxford: Oxford University Press

Sheffield, J.P. Open Peer Review: Collective Intelligence as a Framework for Theorizing Approaches to Peer Review in the Humanities. *NANO: New American Notes Online* 3 (2013).

Shon, J., Marshall, J., & Musen, M. A. (2000). The impact of displayed awards on the credibility and retention of web site information. In *Proceedings of the AMIA Symposium* (p. 794). American Medical Informatics Association.

Sillence, E., Hardy, C., & Briggs, P. (2013, May). Why don't we trust health websites that help us help each other?: an analysis of online peer-to-peer healthcare. In *Proceedings of the 5th Annual ACM Web Science Conference* (pp. 396-404). ACM.

Singer, P., Flöck, F., Meinhart, C., Zeitfogel, E., & Strohmaier, M. (2014, April). Evolution of reddit: from the front page of the Internet to a self-referential community?. In *Proceedings of the 23rd International Conference on World Wide Web* (pp. 517-522). ACM.

Smith, C. (2016). *60 Amazing Reddit Statistics* (November 2016).  
www.expandedramblings.com; last accessed 12 June 2017

Smith, J. B. (1994). *Collective intelligence in computer-based collaboration*. CRC Press.

statista.com (2016) *Number of unique visitors to Reddit from July 2012 to April 2016 (in millions)*. August, 2016. <https://www.statista.com/statistics/443332/reddit-monthly-visitors/>; last accessed 12 June 2017

Stoddard, G. (2015a, April). Popularity Dynamics and Intrinsic Quality in Reddit and Hacker News. In *ICWSM* (pp. 416-425).

Stoddard, G. (2015b, May). Popularity and quality in social news aggregators: A study of reddit and hacker news. In *Proceedings of the 24th International Conference on World Wide Web* (pp. 815-818). ACM.

Sturcke, J., (2009), Swine Flu Fears Boost NHS Direct Queries, *the Guardian*, 28 April, 2009 <https://www.theguardian.com/world/2009/apr/28/swine-flu-nhs-direct-advice>

Suarez-Almazor, M. E., Kendall, C. J., & Dorgan, M. (2001). Surfing the Net – information on the World Wide Web for persons with arthritis: patient empowerment or patient deceit?. *The Journal of Rheumatology*, 28(1), 185-191.

Surowiecki, J. (2005). *The Wisdom of Crowds*. Anchor.

Szokan, N., (Feb, 2011) Internet at 2<sup>nd</sup> Opinion. *The Washington Post* 6 Feb 2011

Tan, C., Niculae, V., Danescu-Niculescu-Mizil, C., & Lee, L. (2016, April). Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In *Proceedings of the 25th International Conference on World Wide Web* (pp. 613-624). International World Wide Web Conferences Steering Committee.

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches* (Vol. 46). Sage.

Taubenberger, J. K., & Morens, D. M. (2006). 1918 Influenza: the Mother of All Pandemics. *Emerging Infectious Disease Journal*. 12 (1). CDC

Terrace, R., Singer, C., & Rosen, R. (2001). Jump-starting digital health. *Jupiter Research Center*.

Thaler, R. H., & Sunstein, C. R. Nudge (2008). *Improving Decisions About Health, Wealth, and Happiness*. New Haven: Yale University Press; 2008.

Tomasello, M. (2009a). *Why we cooperate*. MIT press.

Tomasello, M. (2009b). *The cultural origins of human cognition*. Harvard University Press.

Tomasello, M. (2014). *A natural history of human thinking*. Harvard University Press.

Turoff, M. (2002). Past and future emergency response information systems. *Communications of the ACM*, 45(4), 29-32.

van der Vaart, R., van Deursen, A. J., Drossaert, C. H., Taal, E., van Dijk, J. A., & van de Laar, M. A. (2011). Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. *Journal of Medical Internet Research*, 13(4), e86.

Van Mieghem, P. (2011). The N-intertwined SIS epidemic network model. *Computing*, 93(2-4), 147-169.

van Mierlo T. (2014) The 1% Rule in Four Digital Health Social Networks: An Observational Study. *Journal of Medical Internet Research*, 16(2):e33

Venkatesh, V., & Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision sciences*, 39(2), 273-315.

Venkatesan, S., Abdelhamid, M., Monteiro, J., & Sharman, R. (2016). Internet as Preferred Health Information Source: An Information Foraging Perspective.

Venkatesh, V., & Davis, F. D. (1996). A model of the antecedents of perceived ease of use: Development and test. *Decision sciences*, 27(3), 451-481.

- Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science*, 46(2), 186-204.
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, 425-478.
- Wadley, G., Smith, W., Ploderer, B., Pearce, J., Webber, S., Whooley, M., & Borland, R. (2014, December). What people talk about when they talk about quitting. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: the Future of Design* (pp. 388-391). ACM.
- Wang, Z., Walther, J. B., Pingree, S., & Hawkins, R. P. (2008). Health information, credibility, homophily, and influence via the Internet: Web sites versus discussion groups. *Health communication*, 23(4), 358-368.
- Ward, L. F. (1906). *Applied sociology* (p. 384). Boston: Ginn.
- Wathen, C. N., & Burkell, J. (2002). Believe it or not: Factors influencing credibility on the Web. *Journal of the American society for information science and technology*, 53(2), 134-144.
- Weintraub, K., (2014), When It Comes to Ebola, What Does Quarantine Really Mean? *National Geographic*, October 28, 2014.
- Weninger, T., Zhu, X. A., & Han, J. (2013, August). An exploration of discussion threads in social news sites: A case study of the reddit community.

In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 579-583). ACM.

Weninger, T. (2014). An exploration of submissions and discussions in social news: Mining collective intelligence of Reddit. *Social Network Analysis and Mining*, 4(1), 1-19.

Weninger, T., Johnston, T. J., & Glenski, M. (2015, August). Random voting effects in social-digital spaces: A case study of reddit post submissions. In *Proceedings of the 26th ACM conference on hypertext & social media* (pp. 293-297). ACM.

Whitelaw, N., Bhattacharya, S., McLernon, D., & Black, M. (2014). Internet information on birth options after caesarean compared to the RCOG patient information leaflet; a web survey. *BMC pregnancy and childbirth*, 14(1), 361.

Wilson, P. M. (2001). A policy analysis of the Expert Patient in the United Kingdom: self-care as an expression of pastoral power?. *Health & social care in the community*, 9(3), 134-142.

Wired (2006). Wired Digital Acquires Reddit. *Wired News Report*, 31 October 2006. <http://www.wired.com/2006/10/wired-digital-acquires-reddit/>; last accessed 12 June 2017

Woolley, A. W., Aggarwal, I., & Malone, T. W. (2015). Collective intelligence and group performance. *Current Directions in Psychological Science*, 24(6), 420-424.

World Health Organization (2008). *Outbreak Communication Planning Guide*, World Health Organization. 2008

World Health Organization. Global Influenza Programme. (2009). *Pandemic influenza preparedness and response: a WHO guidance document*. World Health Organization.

World Health Organization. (2016a) International Health Regulations (2005) Third Edition 2016. World Health Organization. 2016

World Health Organisation (2016b), Statement on the 9<sup>th</sup> meeting of the IHR Emergency Committee regarding the Ebola outbreak in West Africa, 29 March 2016. World Health Organisation, 2016.

Wray, R., & Jupka, K. (2004). What does the public want to know in the event of a terrorist attack using plague? *Biosecurity and bioterrorism: biodefense strategy, practice, and science*, 2(3), 208-215.

Wray, R. J., Becker, S. M., Henderson, N., Glik, D., Jupka, K., Middleton, S., ... & Mitchell, E. W. (2008). Communicating with the public about emerging health threats: lessons from the pre-event message development project. *American Journal of Public Health*, 98(12), 2214-2222.



**The Role of Online  
Discussion Forums During a  
Public Health Emergency**

**Jennifer Ann Cole**

**Appendices to the  
Thesis submitted for the degree of  
Doctor of Philosophy**



**Royal Holloway  
University of London**

**2017**



## APPENDICES

---

- Appendix I**      *Questions for People Stationed in Ebola Affected Regions of the World* (1 page) from PHASE I: Interviews with Ebola Witnesses. Interview guide
- Appendix II**      *Quality of Online Health Information: Discussion Threads for Assessment* (7 pages) from PHASE II-A: Investigation of information quality in online health forums (pilot study). Instructions for participants
- Appendix III**      *Discussion Threads for Assessment: Results* (4 pages), from PHASE II-A: Investigation of information quality in online health forums (pilot study). Results data.
- Appendix IV**      *Health Information Seeking During the 2014-16 Ebola Outbreak: A Study on the Quality of Health Information* (17 pages) from PHASE II-B: Investigation of information quality in r/ebola. Instructions for survey participants.
- Appendix V**      *Investigation of information quality in r/ebola: Results data* (2 pages) from PHASE II-B Investigation of information quality in r/ebola. Results data.
- Appendix VI**      *Questions for moderators of [the reddit xxx forum]:* (4 pages) from PHASE III: Observation of reddit and interviews with moderators. Interview guide.



# **Appendix I**

## **Questions for People Stationed in Ebola-Affected Regions of the World**

### PHASE I: Interviews with Ebola Witnesses

#### Interview guide

(1 page)

Appendix I contains the semi-structured interview guide prepared for the interviews with the West Africa Group undertaken for Phase I: Interviews with Ebola Witnesses

It is presented on the following pages exactly as it was used during the interviews.



## **APPENDIX I: PHASE I – INTERVIEWS WITH EBOLA WITNESSES. INTERVIEW GUIDE**

---

### **QUESTIONS FOR PEOPLE STATIONED IN EBOLA-AFFECTED REGIONS OF THE WORLD**

**Semi-structured questions intended as prompts for face-to-face or telephone/Skype interviews. Interviewees will not be shown the questions prior to the interview, but will be given the option of expanding on some of their answers afterwards if required.**

1. What was the first indication you had that there was a serious infectious disease near your location?
2. Who/where did this information come from? (person/organisation and platform)
3. What were your first thoughts?
4. What did you most want to know?
5. How aware were you of Ebola before this, and how did this affect your reaction to the news?
6. Did you feel that the information you wanted was easily available?
7. Did you feel the information you received was honest and gave a full account of the current situation?
8. How did you keep up-to-date with the situation?
9. How well do you feel you understand how Ebola is transmitted, what symptoms it causes, and how much of a threat it was to you?
10. In a similar situation in future, how do you think information requirements might be handled differently?



# **Appendix II**

## **Quality of Online Health Information: Discussion Forum Threads for Assessment**

PHASE II-A: Investigation of information quality in online health forums  
(pilot study).

Instructions for survey participants  
(7 pages)

Appendix II contains the instructions on how to complete a series of online surveys. It was given to the doctors and members of the public who participated in Phase II-A of the study, Investigation of information quality in online health forums.

The survey contains links to the online information they were asked to assess, and to the survey where they could make their assessments. All links are still live. The instructions are presented here exactly as they were seen by the survey respondents.



**APPENDIX II: PHASE II-A – INVESTIGATION OF INFORMATION  
QUALITY IN ONLINE HEALTH FORUMS (PILOT STUDY)**

---

**QUALITY OF ONLINE HEALTH INFORMATION: DISCUSSION FORUM  
THREADS FOR ASSESSMENT**

**QR1: Diabetes: I am going to a party for the first time since being diagnosed [with Type 1 diabetes]. Does anyone have advice?**

Link to discussion:

[http://www.reddit.com/r/diabetes/comments/2vl8h6/i\\_am\\_going\\_to\\_a\\_party\\_for\\_the\\_first\\_time\\_since\\_i/](http://www.reddit.com/r/diabetes/comments/2vl8h6/i_am_going_to_a_party_for_the_first_time_since_i/)

Link to survey questionnaire:

<http://www.smartsurvey.co.uk/s/WG62L/>

**QR2. Diabetes: Advice for exercise and midnight lows?**

Link to discussion:

[http://www.reddit.com/r/diabetes/comments/2vdx1v/advice\\_for\\_exercise\\_and\\_midnight\\_lows/](http://www.reddit.com/r/diabetes/comments/2vdx1v/advice_for_exercise_and_midnight_lows/)

Link to survey questionnaire:

<http://www.smartsurvey.co.uk/s/HWT4I/>

**QM3. Diabetes: Are anger outbursts normal with Diabetes?**

Link to discussion:

[http://www.mumsnet.com/Talk/general\\_health/a1258323-Are-anger-outbursts-normal-with-diabetes](http://www.mumsnet.com/Talk/general_health/a1258323-Are-anger-outbursts-normal-with-diabetes)

Link to survey: <http://www.smartsurvey.co.uk/s/ZQM1Y/>

**QM4. Diabetes: Signs of diabetes or paranoid Mummy?**

Link to discussion:

[http://www.mumsnet.com/Talk/childrens\\_health/a1298036-Signs-of-diabetes-or-paranoid-mummy](http://www.mumsnet.com/Talk/childrens_health/a1298036-Signs-of-diabetes-or-paranoid-mummy)

Link to survey questionnaire: <http://www.smartsurvey.co.uk/s/XQKY7/>

**QP5. Diabetes: Does it take longer to get over a cold if you have Type 2 Diabetes?**

Link to discussion:

<http://www.patient.co.uk/forums/discuss/common-cold-360470>

Link to survey questionnaire: <http://www.smartsurvey.co.uk/s/RSC2C/>

**QP6. Diabetes: Can this be Diabetes?**

Link to discussion:

<http://www.patient.co.uk/forums/discuss/can-this-be-diabetes--339377>

Link to survey questionnaire:

<http://www.smartsurvey.co.uk/s/ZMOE3/>

**QP7. Do I have Type 1 Diabetes?**

Link to discussion:

<http://www.patient.co.uk/forums/discuss/do-i-have-type-one-diabetes--338989>

Link to survey questionnaire:

<http://www.smartsurvey.co.uk/s/Q6WS3/>

**QP8. Diabetes: Advice please?**

Link to discussion:

<http://www.patient.co.uk/forums/discuss/advice-please-299215>

Link to survey questionnaire: <http://www.smartsurvey.co.uk/s/0V6AS/>

**QR9. HIV/AIDS: I found out I had HIV but not clear about the stage**

Link to the discussion:

[http://www.reddit.com/r/hiv aids/comments/2gdck8/i\\_found\\_out\\_i\\_have\\_hiv\\_but\\_not\\_clear\\_about\\_the/](http://www.reddit.com/r/hiv aids/comments/2gdck8/i_found_out_i_have_hiv_but_not_clear_about_the/)

Link to the survey questionnaire: <http://www.smartsurvey.co.uk/s/Z068l/>

**QR10. HIV/AIDS: HIV and depression?**

Link to discussion:

[http://www.reddit.com/r/hiv aids/comments/2u2zgp/hiv\\_and\\_depression/](http://www.reddit.com/r/hiv aids/comments/2u2zgp/hiv_and_depression/)

Link to survey questionnaire: <http://www.smartsurvey.co.uk/s/I0KY3/>

**QR11. HIV/AIDS: HIV and personal fitness?**

Link to discussion:

[http://www.reddit.com/r/hiv aids/comments/228l04/question\\_about\\_hiv\\_and\\_personal\\_fitness/](http://www.reddit.com/r/hiv aids/comments/228l04/question_about_hiv_and_personal_fitness/)

Link to survey questionnaire:

<http://www.smartsurvey.co.uk/s/ZPJ2W/>

**QR12. HIV/AIDS: HIV FAQs**

Link to discussion:

[http://www.reddit.com/r/hiv aids/comments/1b5oeh/faq\\_are\\_you\\_worried\\_about\\_exposure\\_risk\\_testing/](http://www.reddit.com/r/hiv aids/comments/1b5oeh/faq_are_you_worried_about_exposure_risk_testing/)

Link to survey questionnaire: <http://www.smartsurvey.co.uk/s/OZPA1/>

**QM13: HIV/AIDS: Babysitter has just announced he's HIV. Should I worry?**

Link to discussion:

[http://www.mumsnet.com/Talk/childminders\\_nannies\\_au\\_pairs\\_etc/a499271-babysitter-and-friend-has-just-announced-he-s-hiv-positive](http://www.mumsnet.com/Talk/childminders_nannies_au_pairs_etc/a499271-babysitter-and-friend-has-just-announced-he-s-hiv-positive)

Link to survey questionnaire: <http://www.smartsurvey.co.uk/s/JMRGM/>

**QM14: HIV/AIDS: Question about HIV (risk from partner)?**

Link to discussion:

[http://www.mumsnet.com/Talk/general\\_health/a85259-question-about-hiv](http://www.mumsnet.com/Talk/general_health/a85259-question-about-hiv)

Link to survey questionnaire: <http://www.smartsurvey.co.uk/s/CA27W/>

**QM15: HIV/AIDS: Question about children and HIV**

Link to discussion:

[http://www.mumsnet.com/Talk/general\\_health/a543209-children-and-hiv](http://www.mumsnet.com/Talk/general_health/a543209-children-and-hiv)

Link to survey questionnaire:

<http://www.smartsurvey.co.uk/s/38OQ1/>

**QP16: HIV/AIDS: HIV Question**

Link to discussion:

<http://www.patient.co.uk/forums/discuss/hiv-question-358541>

Link to survey questionnaire: <http://www.smartsurvey.co.uk/s/1XEHL/>

**QP17: HIV/AIDS: Can a possible HIV infection cause intestinal yeast after just 4 months?**

Link to discussion: <http://www.patient.co.uk/forums/discuss/can-a-possible-hiv-infection-cause-intestinal-yeast-after-only-4-months--337964>

Link to survey questionnaire: <http://www.smartsurvey.co.uk/s/JLWOW/>

**QR18. Chickenpox: Is this chickenpox? Help!**

Link to discussion:

[http://www.reddit.com/r/reddit.com/comments/bso73/is\\_this\\_chickenpox\\_help/](http://www.reddit.com/r/reddit.com/comments/bso73/is_this_chickenpox_help/)

Link to survey questionnaire:

<http://www.smartsurvey.co.uk/s/LEU3X/>

**QR19. Chickenpox: Did you give your child the chickenpox vaccine?**

Link to discussion:

[http://www.reddit.com/r/Parenting/comments/1aiqlh/did\\_you\\_give\\_your\\_child\\_the\\_chickenpox\\_vaccine/](http://www.reddit.com/r/Parenting/comments/1aiqlh/did_you_give_your_child_the_chickenpox_vaccine/)

Link to survey questionnaire:

<http://www.smartsurvey.co.uk/s/BJX1T/>

**QR20. Chickenpox: Why are some diseases such as chickenpox more dangerous to adults than children?**

Link to discussion:

[http://www.reddit.com/r/askscience/comments/r3ml9/why\\_are\\_some\\_diseases\\_such\\_as\\_chickenpox\\_more/](http://www.reddit.com/r/askscience/comments/r3ml9/why_are_some_diseases_such_as_chickenpox_more/)

Link to survey questionnaire: <http://www.smartsurvey.co.uk/s/ZFQXD/>

**QM21. Chickenpox: is 5 months too young to expose?**

Link to discussion:

<http://www.mumsnet.com/Talk/parenting/a1200011-Chicken-pox-is-5-months-too-young-to-expose>

Link to survey questionnaire: <http://www.smartsurvey.co.uk/s/U4PXN/>

**QM22. Chickenpox: Have you given your child the chickenpox vaccine?**

Link to discussion:

<http://www.mumsnet.com/Talk/vaccinations/a1552540-have-you-given-your-child-the-chicken-pox-vaccine>

Link to survey questionnaire: <http://www.smartsurvey.co.uk/s/3ZCUT/>

**QM23. Chickenpox: Is it normal to be so very ill with Chickenpox?**

Link to discussion:

[http://www.mumsnet.com/Talk/childrens\\_health/a1916107-Is-it-normal-to-be-so-very-ill-with-chicken-pox](http://www.mumsnet.com/Talk/childrens_health/a1916107-Is-it-normal-to-be-so-very-ill-with-chicken-pox)

Link to survey questionnaire: <http://www.smartsurvey.co.uk/s/RAF72/>

**QP24. Chickenpox: Should my toddler get the chickenpox vaccine?**

Link to discussion:

<http://www.patient.co.uk/forums/discuss/should-my-toddler-get-the-chickenpox-vaccine--258080>

Link to survey questionnaire: <http://www.smartsurvey.co.uk/s/BUXCH/>

**QP25. Chickenpox: Strange symptom with chickenpox**

Link to discussion:

<http://www.patient.co.uk/forums/discuss/strange-symptom-with-chicken-pox-257950>

Link to survey questionnaire:

<http://www.smartsurvey.co.uk/s/6FZVT/>



# **Appendix III**

## **Discussion Threads for Assessment: Results**

PHASE II-A: Investigation of information quality in online health forums  
(pilot study)  
Results data.  
(4 pages)

Appendix III contains the data table of results from the survey undertaken by the participants in Phase II-A of the study, Investigation of information quality in online health forums.



**APPENDIX III: PHASE II:A – DISCUSSION THREADS FOR ASSESSMENT: RESULTS**

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	1	2	
			1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	
DIABETES			Accurate					Complete					Sensible					Appropriate					Useful					
Reddit	Q1. First party since being diagnosed, need advice?	M1	■					■						■				■					■					
		M2	■					■						■					■					■				
		P9	■						■						■					■					■			
	Q2. Advice for exercise and midnight lows?	M1	■						■						■				■						■			
M2		■						■						■					■					■				
Mumsnet	Q3. Are anger outbursts normal with diabetes?	M1	■						■					■					■					■				
		M2			■					■					■					■					■			
		P1			■					■					■					■					■			
		P9	■							■			■			■					■				■			
Q4. Signs of diabetes or paranoid Mummy?	M1	■							■					■					■					■				
	M2	■							■					■					■					■				
	P1	■							■					■					■					■				
Patientco.uk	Q5. Longer to get over a cold with diabetes?	M1										■							■					■				
		M3			■									■						■					■			
		P2			■					■						■					■					■		
		P3				■									■						■					■		
Q6. Can this be Diabetes?	M1	■							■					■					■					■				
	P2			■					■						■					■					■			
Q7. Do I have Type 1 Diabetes?	M1	■							■					■					■					■				
	M2	■							■					■					■					■				
	P1	■							■					■					■					■				
Q8. Diabetes: Advice please?	M1	■							■					■					■					■				
	P2			■					■						■					■					■			

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	1	2		
			1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5		
<b>HIV/AIDS</b>			Accurate					Complete					Sensible					Appropriate					Useful						
Reddit	Q9. I found out I had HIV not clear about the stage	M1	■						■					■					■					■					
		M4	■						■						■					■					■				
	Q.10 HIV and depression	M1		■								■				■						■						■	
		M4		■							■					■						■					■		
	Q11. Question about HIV and personal fitness	M1		■						■					■							■					■		
		M4		■						■					■							■					■		
	Q12. FAQ: Worried? Risk, testing and anxiety	M1	■							■					■							■					■		
		M4	■							■					■							■					■		
Mumsnet	Q13. Babysitter has just announced he's HIV positive	M1		■					■					■							■					■			
		M4		■						■					■						■					■			
	Q.14. Question about HIV (and partner)	M1		■						■					■							■					■		
		M4		■						■					■							■					■		
	Q15. Children and HIV	M1		■						■					■							■					■		
		M4		■						■					■							■					■		
Patient.co.uk	Q16. HIV question	M1		■					■					■							■					■			
		M4		■						■					■						■					■			
	Q17. HIV infection: intestinal yeast after 4 months?	M1		■						■					■							■					■		
		M4			■						■					■						■					■		

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	1	2
			1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	<b>CHICKENPOX</b>		Accurate					Complete					Sensible					Appropriate					Useful				
Reddit	Q18. Is this chickenpox? Help!!	M1					■					■					■					■					■
		M4			■						■					■					■					■	
		M5				■					■					■					■					■	
		M6					■				■					■					■					■	
		P4			■						■					■					■					■	
		P5					■				■					■					■					■	
		P6				■					■					■					■					■	
Mumsnet	Q19. Did you give your child the chickenpox vaccine?	M1		■							■					■					■					■	
		M4		■							■					■					■					■	
		M5		■							■					■					■					■	
		M6				■					■					■					■					■	
		P4				■					■					■					■					■	
		P5			■						■					■					■					■	
		P6				■					■					■					■					■	
Mumsnet	Q20. Chickenpox: Why more dangerous to adults?	M1				■					■					■					■					■	
		M5				■					■					■					■					■	
		M7				■					■					■					■					■	
	Q21. Chickenpox: is 5 months too young to expose?	M1				■					■					■					■					■	
	M4			■						■					■					■					■		
	M5					■				■					■					■					■		
	P4				■					■					■					■					■		
	P7				■					■					■					■					■		
	P8					■				■					■					■					■		

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	1	2	
			1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	
		<b>CHICKENPOX (Continued)</b>	Accurate					Complete					Sensible					Appropriate					Useful					
Patient.co.uk	Q22. Has your child had the chickenpox vaccine?	M1				■						■					■										■	
		M4	■							■				■						■							■	
		M5				■					■							■									■	
		P6				■						■						■									■	
		P7				■					■							■									■	
	Q23. Is it normal to be so very ill with chickenpox?	M1					■						■					■									■	
		M4	■								■							■									■	
		M8	■								■			■					■								■	
		P4				■						■						■									■	
	Q24. Should toddler get the chickenpox vaccine?	M1	■								■							■									■	
		M4	■								■							■									■	
		M5	■								■							■									■	
Q25. Strange symptom with chickenpox	M1					■						■					■									■		
	M4				■						■						■									■		
	M5				■						■						■									■		

## **Appendix IV**

### **Health Information Seeking During the 2014-16 Ebola Outbreak: A Study on the Quality of Health Information**

PHASE II-B: Investigation of information quality in r/ebola.

Instructions for survey participants.

(17 pages)

Appendix IV contains the survey given to respondents who participated in Phase II-B of the study, Investigation of information quality in r/ebola. The survey was completed by doctors at the Royal Society of Medicine Military Medicine conference held in September 2016, and by doctors from the West Kent Medico Chirurgical Society at a meeting held in November 2016.

The survey presented here is exactly as it was seen by the doctors.



# **Health Information Seeking During the 2014-16 Ebola Outbreak A Study on the Quality of Health Information**

**Jennifer Cole**  
PhD Candidate, Royal Holloway University of London  
Jennifer.Cole.2013@live.rhul.ac.uk

## **Health Information Seeking During the 2014-16 Ebola Outbreak: Study on the Quality of Health Information**

---

Thank you for agreeing to take part in this survey. It is part of a study on health information during public health emergencies being carried out under a PhD programme at Royal Holloway, University of London, funded by a Reid Scholarship in Health, the Human Body and Behaviour (H2B2).

On the following pages, you will find a series of questions which were asked by mainly US-based health information seekers with regard to the risk from Ebola to their own health, during September and October 2014. At this point in time, the Ebola outbreak was at its height in West Africa. Infected US aid workers were being flown back to the US for treatment and there were a handful of cases diagnosed on US soil, which was making US citizens nervous that there may be further cases in the US.

Each question is presented on a separate page. Following each question, we present four different answers that were given in response to it.

We would like you to rate each of the answers as **good**, **bad** or **neutral** in accordance with the criteria set out on the following page.

For each set of answers to a particular question, any score pattern can be given, i.e. you may consider all the answers to one question to be good, or all the answers to be bad, or for the answers to cover a range in which some are good, some are bad, or some are neutral. Please do not assume that each answer set contains examples of good, neutral and bad answers.

Please score each answer in absolute terms i.e. good or bad in its own right, rather than better or worse than the other answers with which it is presented. Some questions will have received better answers overall than others.

At the end of each question and answers set, there is a box to explain why you have rated the questions in the way you have. Do not feel obliged to add any information here, but if you do, please add as much or as little as you like. Please also use this box to list any information you think should have been included in an answer for it to be considered 'good', but which was missing from any of the answers offered.

If you would like more information, please contact Jennifer Cole at [Jennifer.Cole.2013@live.rhul.ac.uk](mailto:Jennifer.Cole.2013@live.rhul.ac.uk)

## Criteria for marking answers

---

### **Good - information should be marked as good if it displays some or all of the following criteria:**

- It gives information which is accurate, and is in line with what would be expected from a qualified medical or public health practitioner
- It gives information which a qualified professional would consider to be sensible and appropriate in light of the question asked
- It gives information which is unbiased, and does not present one view as the only available option if more than one is available

### **Neutral – information should be marked as neutral if it displays some or all of the following criteria:**

- It gives information which, while it does not contain factual inaccuracies, is unlikely to add anything to the questioner's existing knowledge or understanding of their situation.
- It gives information which a qualified medical or public health practitioner would not necessarily consider to have any relevance to the question that has been asked, but which does not necessarily provide inappropriate information.
- It provides no noticeable value to the discussion, but neither is it in any way objectionable.
- It contains a mix of both good and bad information which, when taken together, means that it cannot be considered 'good' or 'bad'

### **Bad – information should be marked as bad if it displays some or all of the following criteria:**

- It gives information which contains factual or scientific errors and inaccuracies
- It gives information which a qualified medical or public health practitioner would not consider to be a sensible and appropriate response to the question posed
- There is an obvious piece of advice that should have been given, but which is not part of the answer provided.
- It gives information which is heavily biased towards one course of action, and discourages the questioner from taking others (particularly ones which a medical or public health practitioner would consider to be more appropriate)
- It is racist, sexist or in any way abusive
- It is overly flippant or rude to such an extent that it is likely to be unhelpful even if the information it contains is technically accurate and correct.

**Question 1:**

Why are so many health professionals catching Ebola? While I understand the argument that they are of course around the infected more and that this is leading them to catch it, but given the amount of knowledge that they have about it, why are they still catching it? I am curious if they are being misinformed on how it is transmitted or how contagious it is.

Answer 1			Answer 2		
I think that there is a lack of understanding, or of acceptance of the potency of droplet and aerosol transmission (not airborne, I know) both near the patient and near their discarded belongings and waste. I've read that during the worst part of the infection, they can produce bagfulls of waste per day.			Count how many times you touch your face in the next 15 minutes. Imagine now trying to be super diligent for weeks at a time, in pretty horrible conditions, in hot humid conditions. Take a look at some of the pictures from the recent <i>NY Times</i> article about conditions in the Ebola hospitals in Liberia.		
Good	Neutral	Bad	Good	Neutral	Bad
Answer 3			Answer 4		
At some point you have to change out of your suit. If you miss a step or do things in a wrong order, you may just have contaminated yourself. It only takes one slip or one touch to the lips and you've transmitted the virus. It's very hard to maintain 100% vigilance and 100% compliance all day every day, even if you are properly equipped and know what you are doing.			In the highly affected areas in west Africa, personnel are spread thin with few supplies. Ebola treatment centers are mostly just places patients go to be isolated from the world and receive barely any real treatments. These places are very poorly maintained and because of this, the healthcare workers in them are very likely to be exposed on surfaces that are not clean or from handling patients with sub-par or nonexistent protective equipment. Outside of that, there is a lack of training too. Handling patients that require specific protective procedures are not things that every healthcare worker has much experience with. They could be using all the gear they're instructed to, but a mistake in taking the equipment off after handling the patient or handling things such as soiled linens can lead to exposure.		
Good	Neutral	Bad	Good	Neutral	Bad

Please use the box below to give any comments you think are useful for explaining the scores you gave:

**Question 2:**

Are plumbers at risk? I work for a plumbing company in the Dallas area. What are the risks for plumbers servicing the Texas Health Presbyterian Hospital [Where Ebola patient Thomas Eric Duncan was treated and died]. Can they contract Ebola from working with the waste systems? I imagine they can but I haven't heard anyone talking about this. Can plumbers or health professionals give any insight to their risk?

Answer 1			Answer 2		
<p>Totally not an expert and mostly working with bro logic here, but I would imagine that it would be possible, but unlikely for you to be at any measurably increased risk than any other average worker in the same hospital. Yes, you have to deal with human waste, which would bring you at an increased chance of exposure IF you encounter a particle of poo that has Ebola. Then, it would need to enter your mouth, nose or eyes. My understanding, however, is that you would need to encounter an actual piece of poo from an infected person. Once again... I am an armchair speculator, but that is my 2 cents.</p>			<p>Ebola isn't waterborne - the virus dies instantly in bleach, only lasts a few minutes in fresh water, but I can imagine it lasting longer in solids of various sorts. Out of curiosity, what are plumbers trained to do to protect themselves against waterborne pathogens when working on waste systems? (Observed behavior, single sample, plumber hired to deal with main drain blockage in my single-family house: only PPE is rubber gloves, doesn't seem to worry too much about stuff dripping on their skin or clothes or basement floor).</p>		
Good	Neutral	Bad	Good	Neutral	Bad
Answer 3			Answer 4		
<p>Plumbers only know two things:</p> <ol style="list-style-type: none"> <li>1. shit flows downhill</li> <li>2. payday is Friday</li> </ol>			<p>Untouchables are at risk:</p> 		
Good	Neutral	Bad	Good	Neutral	Bad

Please use the box below to give any comments you think are useful for explaining the scores you gave:

**Question 3:**

How effective is alcohol-based sanitizer? Does it even kill Ebola viruses, and if not then why are health officials telling the public to use it? If it does, then we need to be passing out bottles to everyone we can over in West Africa. Also, if it doesn't, they need to come up with a bleach-based sanitizer to pass out. Just an idea.

Answer 1			Answer 2		
It's more of a bleach/water mix. The public needs to be using sanitizer like this instead of an alcohol sanitizer.			The US CDC and WHO state waterless <b>alcohol-based hand sanitizer</b> may be used as long as hands are physically clean/not visibly soiled. See: <a href="https://www.internationalsos.com/ebola/index.cfm?content_id=410&amp;">https://www.internationalsos.com/ebola/index.cfm?content_id=410&amp;</a> Susceptibility to disinfectants: Ebolavirus is susceptible to 3% acetic acid, 1% glutaraldehyde, <b>alcohol-based products</b> , and dilutions (1:10-1:100 for ≥10 minutes) of 5.25% household bleach (sodium hypochlorite), and calcium hypochlorite (bleach powder) [references to these provided]. The WHO recommendations for cleaning up spills of blood or body fluids suggest flooding the area with a 1:10 dilutions of 5.25% household bleach for 10 minutes for surfaces that can tolerate stronger bleach solutions (e.g., cement, metal). For surfaces that may corrode or discolour, they recommend careful cleaning to remove visible stains, followed by contact with a 1:100 dilution of 5.25% household bleach for more than 10 minutes. See: <a href="http://www.phac-aspc.gc.ca/lab-bio/res/psds-ftss/ebola-eng.php">http://www.phac-aspc.gc.ca/lab-bio/res/psds-ftss/ebola-eng.php</a>		
Good	Neutral	Bad	Good	Neutral	Bad
Answer 3			Answer 4		
Pure alcohol would evaporate fast leaving not much of a fire hazard. If alcohol is that effective they could easily spray workers down with pure alcohol and remain safe, and we might not see as many health worker infections.			There is a reason the people dealing with the patients are using a chlorine solution and not hand sanitizer.		
Good	Neutral	Bad	Good	Neutral	Bad

Please use the box below to give any comments you think are useful for explaining the scores you gave:

**Question 4:**

Why is Ebola more dangerous than the SARS epidemic? SARS is far more infectious. It has an  $R_0$  of 4 vs Ebola with 2 and it's airborne. By 2003, SARS had infected over 8000 people before it was stopped. Why were we able to contain SARS, but Ebola will be so much more difficult?

Answer 1			Answer 2		
<p>It's not [more dangerous than SARS]. SARS was far more dangerous and caused a far greater amount of panic at the time than I see now. I remember the wall-to-wall coverage, the stock market drops, the predictions of certain doom. One critical difference was the response of the Chinese, when they had a few hundred cases in Beijing, they started building a field hospital for 10,000. They anticipated the exponential growth and got on top of it. They shut down Beijing, for crissakes! We (the world as a whole) dodged a bullet with SARS, I hope we can dodge this one too, and the one after that, and so on. Humanity is a remarkably resilient species, we aren't going down without a fight.</p>			<p>An important difference is Ebola is still happening. It's like looking outside an airliner window and seeing an engine burst into flames and asking "why is this more dangerous than the other case when the same thing happened and the aircraft safely landed?" People get this insane idea that SARS wasn't dangerous because it turned out okay. If you play Russian roulette, do you afterwards say, "what a fool I was, there was no bullet in the barrel so I had no reason to be afraid?" We can't actually see the future. We make educated guesses. Some of the guesses about Ebola are really bad. What compels us to act is the very uncertainty. We don't know how bad it will be: we've never been in this situation.</p>		
Good	Neutral	Bad	Good	Neutral	Bad
Answer 3			Answer 4		
<p>the SARS epidemic had a lower fatality rate. Much lower. 8,096 cases and 774 deaths in the 2002-2003 outbreak. We're about to break 5,000 reported deaths of 10K+ cases. And this is just what's reported. Ebola is spreading slower, but it's still spreading fast. It's also spreading in areas where the public health infrastructure is insufficient to provide adequate care to independently contain infection rates. Even if we get vaccines and medications on the ground soon, after this hits a tipping point, Ebola won't be something West Africa will be able to eradicate at least for another a few years to a few decades.</p>			<p>There are two variables: <i>contagious</i> vs <i>infectious</i>, which while similar, mean different things. (I hope I'm not mixing this up but) <i>contagious</i> refers to how quickly a disease can spread. So an airborne pathogen is much more <i>contagious</i> than a nonairborne pathogen (ie, SARS is more contagious than Ebola). <i>Infectious</i> refers to how much load of an organism must you acquire before becoming infected. In this regard, Ebola is more <i>infectious</i> than SARS (correct me if wrong). The mortality rate is also different for SARS and Ebola. You can get Ebola with a single drop of infected fluid whereas you may or may not get symptomatic SARS even if someone coughed right in front of you. Ebola is much harder to contain. You don't need the high level of PPE/decontamination for SARS as you do for Ebola. For SARS, you just need to wear a mask and wash your hands.</p>		
Good	Neutral	Bad	Good	Neutral	Bad

Please use the box below to give any comments you think are useful for explaining the scores you gave:

**Question 5:**

How long does Ebola survive on bills and coins?

Answer 1			Answer 2		
I could see an end to cash transactions in the West.			Filoviruses have been reported capable to survive for weeks in blood and can also survive on contaminated surfaces, particularly at low temperatures (4°C). One study could not recover any Ebolavirus from experimentally contaminated surfaces (plastic, metal or glass) at room temperature, in another, Ebolavirus dried onto glass, polymeric silicone rubber, or painted aluminum alloy is able to survive in the dark for several hours under ambient conditions. When dried in tissue culture media onto glass and stored at 4 °C, <i>Zaire ebolavirus</i> survived for over 50 days. This information is based on experimental findings only and not based on observations in nature. This information is intended to be used to support local risk assessments in a laboratory setting. See: <a href="http://www.phac-aspc.gc.ca/lab-bio/res/psds-ftss/ebola-eng.php">http://www.phac-aspc.gc.ca/lab-bio/res/psds-ftss/ebola-eng.php</a> How long can the Ebola survive outside the body? It varies a lot (temperature, humidity, pH, etc.), but Ebola can survive for one to two days outside the body: <a href="http://www.emsworld.com/article/11616877/cdc-answers-ebola-questions">http://www.emsworld.com/article/11616877/cdc-answers-ebola-questions</a> It's reported that the banks that are still open in the Ebola regions have buckets full of bleach water at their entrances with guards to enforce hand washing. Tellers are wearing latex gloves.		
Good	Neutral	Bad	Good	Neutral	Bad
Answer 3			Answer 4		
Which brings up my other favorite [topic] - Bitcoins. No physical contact. No handing bills. No sliding cards.			Oh geeze, now I have to worry about this. I was flummoxed about not taking a pedicure due to risk.		
Good	Neutral	Bad	Good	Neutral	Bad

Please use the box below to give any comments you think are useful for explaining the scores you gave:

**Question 6:**

Can you get Ebola from an airplane bathroom? I'm flying from Australia to JFK in two days, so I've been researching a lot. Every article states you can only catch Ebola from bodily fluids, which I understand. My worry scenario is that if someone with the virus is contagious and uses the bathroom and urinates on the seat, let's say, then I use the toilet seat, can I then get it? I also have scalp psoriasis, which is pretty much open wounds. If someone sweats on the seat then could I contract it?

Answer 1			Answer 2		
It's important to recognize that people are only infectious when they're symptomatic. How infectious a person is relates directly to how sick that person is (as the amount of virus in their system grows, so do the debilitating symptoms). Therefore, someone who is well enough to travel is quantifiably less infectious than someone who is incredibly ill. Ebola can be destroyed by something as simple as bleach, soap or handwash that contains alcohol. Patrick Sawyer infected the official that collected him at the airport, likely when he vomited in the car with him, and the rest were healthcare workers at the hospital.			Patrick Sawyer was pretty far along with Ebola when he flew into Nigeria [from Liberia] and no one on his flight caught it. The guy from the first part in The Hot Zone book flew when he was pretty far along with the disease as well (may have been Marburg though, can't remember) and no one on that flight caught it. Eric Duncan's family stayed with him in a small apartment with only one bathroom until he was pretty far along with the disease and none of them caught it. I also have a long 23hr flight coming up. I'm no longer worried about it.		
Good	Neutral	Bad	Good	Neutral	Bad
Answer 3			Answer 4		
Wear a hat on your head and carry some kind of sanitizer, I'm sure they make some under 2oz or whatever the limit is on gels/fluids. Or maybe snag some liquor off the cart, just use that on the toilet seat if it looks tainted.			Are you a female? Squat on the toilet – don't sit. You should be doing this anyway – bathroom lavs are gross. Wear a scarf over your head to create a barrier. No one will bat an eye about this. You can purchase pretty scarves at the airport. Wash your hands and carry [hand sanitizer gel]. You'll be okay.		
Good	Neutral	Bad	Good	Neutral	Bad

Please use the box below to give any comments you think are useful for explaining the scores you gave:

**Question 7:**

Will getting a flu shot help?

Answer 1			Answer 2		
<p>Me personally? I'm aware of the risks, but not worried. You have to make that decision for yourself, though. There's a risk of getting off your sofa and walking through the door in the morning (could get hit by a car, someone could stab you) and then there's the risk of staying on your sofa and going nowhere (obesity, health problems, etc). Every choice has risks and opportunity costs, choose the risks that you are comfortable with.</p>			<p>A flu shot lessens the chances of catching the flu and will mean you won't get mistaken for an Ebola patient. If you get the flu, you may get paranoid that you have Ebola, creating anxiety. Your illness will fuel the paranoid of others, if/when there are other Ebola cases. You may infect others with flu, spreading fear and doubt about their health and maybe sending them to hospital too. If you go to hospital with flu, you will be a burden to others and may infect others in the waiting room. You may end up sitting next to a real Ebola patient. Getting a flu shot is an easy way to be part of the solution.</p>		
Good	Neutral	Bad	Good	Neutral	Bad
Answer 3			Answer 4		
<p>Meh, it has potential for long term side effects. I wouldn't get it unless you are around vulnerable people.</p>			<p>I am not a conspiracy theorist, but why the f*ck would I want to get a flu shot? How on Earth does getting a flu shot have anything to do with Ebola? It's like saying 'if you don't want herpes, you should drink Diet Coke'. This is the most retarded thing I have ever read in my life.</p>		
Good	Neutral	Bad	Good	Neutral	Bad

Please use the box below to give any comments you think are useful for explaining the scores you gave:

**Question 8:**

If this strain of Ebola is not airborne, how else could it have infected 20 times more people than any strain of Ebola in the past?

<b>Answer 1</b>			<b>Answer 2</b>		
<p>The Ebola epidemic is not 20 times more infectious than previous outbreaks. The disease is the same; the conditions it has been in are different. Liberia, for example, and its healthcare system was and is woefully inadequate for the population. The civil war also contributes to peoples' mistrust of the government, and there are also insufficient numbers of people in these countries that trust Western medicine and would rather go to a traditional healer. In short, these are impoverished, under-educated people with incredibly dense populations and poor, incompetent governments. Basically, West Africa was a perfect storm for a public health emergency.</p>			<p>It's the first time that the virus has taken a foothold in a dense urban area in a poor country. It is not as contagious as the flu (which is airborne). If it was as contagious as the flu, Ebola would already be all over the world.</p>		
Good	Neutral	Bad	Good	Neutral	Bad
<b>Answer 3</b>			<b>Answer 4</b>		
<p>You are oversimplifying things and exhibiting a false cause fallacy. There are lots of reasons why things exhibit exponential growth. In this case, I think the simplest explanation is the political/economic situation in West Africa. It was only a matter of time before an epidemic (from the virus's point of view) became successful. There's no evidence that a single case has been transmitted via air. There's ample evidence that political, economic, industrial, medical (etc) instability is a vector, so let's start there.</p>			<p>The other outbreaks were in isolated villages which then stopped existing. By the time any of those managed to spread somewhere else, there was already a large medical response. This outbreak started in a more densely packed area (a city) and was able to spread to more places....</p>		
Good	Neutral	Bad	Good	Neutral	Bad

Please use the box below to give any comments you think are useful for explaining the scores you gave:

**Question 9:**

Is Ebola transmittable on a bus? I live in Cleveland, Ohio and take the bus to work. It gets very crowded and everyone is shoulder-to-shoulder. Am I pushing it too far, wanting to quit my job so I will be safe? Or is this a leap in the direction of paranoia?

Answer 1			Answer 2		
<p>I don't know if quitting your job right now would be a reasonable approach. To answer your question: if someone had symptoms of Ebola, or had Ebola and coughs, sneezes, spits, vomits, cries, or (probably) sweats on you, and you get that onto a mucous membrane (eyes, mouth, nose, genitals) or open wound, you will get Ebola. So yes, if you're crammed shoulder-to-shoulder on a bus and this Ebola person vomits within a foot or three of you, there's a pretty decent chance of getting it. However, I would wait and see if there's any cases that start popping up that aren't connected and "contact traced", in your city. That would mean that we've lost some of the threads of infection, which means you don't know how close it is. That would be when I start to make plans to get out of town if needed. Even in Africa, it's only infected ~ 0.6-1% of people in infected countries after 8 months. It is a serious issue, but your whole city isn't going to keel over in a short period of time. Take hygiene precautions, and buy some spare rice.</p>			<p>You are totally paranoid and need to calm down.</p>		
Good	Neutral	Bad	Good	Neutral	Bad
Answer 3			Answer 4		
<p>Don't put your hands in your mouth. Wash your hands as soon as you get to work. Ebola can't enter through your skin, it has to enter through a mucus membrane, and unless someone is profusely vomiting onto your face, the only way it is going to get in your body is on your hands.</p>			<p>Quit your job so someone with a sense of reality can have it.</p>		
Good	Neutral	Bad	Good	Neutral	Bad

Please use the box below to give any comments you think are useful for explaining the scores you gave:

**Question 10:**

I'm an 18-year-old female living in England. I suffer from quite extreme anxiety and hypochondria and reading [about Ebola] has made me begin to panic about the potential for this virus to become a pandemic throughout developed countries. It's getting to the point where I'm obsessing over it on a near-constant basis, although I know this is very irrational. I'm wondering what I should do to prepare in the event that it does spread to the UK. What sort of food/equipment do you think is necessary?

Answer 1			Answer 2		
I'd go with the sensible general purpose stuff. A few weeks of dried/canned food, some cases of bottled water, a few rolls of duct tape, a good flashlight, radio, extra batteries, maybe a solar panel you can charge your cell phone with... That way when there's never an Ebola outbreak in your country, you can tell people it's for hurricanes/winter storms/etc and not look too crazy.			I'm a 23-year-old male living in the US and I also suffer from hypochondria. I'm not about to tell you what you want to hear, so if you're expecting "Canned bread, baked beans, re-breather", you're going to be disappointed. Don't worry about this Ebola outbreak. If you were in Algeria or Libya, I might say it's a good time to get a game plan together. But you're not, and it's not. Ebola is breaking out so massively in Africa because the people there are superstitious and downright medieval in their understanding of basic medicine. Don't worry about Ebola. You're in a first world nation, be calm and enjoy the benefit of not succumbing to whatever jungle fever comes and goes.		
Good	Neutral	Bad	Good	Neutral	Bad
Answer 3			Answer 4		
The odds of this spreading are low, and get lower as a function of the distance from the initial outbreak. If I have to explain this to you, you're dim and will be dead soon from things like car versus pedestrian accidents and drowning in the rain. Ebola isn't your biggest worry.			Anxiety will cripple you and make your life hell before Ebola or any other threat has a chance. Your first priority should be to prepare for the rest of your life and reduce your anxiety level. Remove yourself from whatever triggers your anxiety. Get counseling and adjust your medications.		
Good	Neutral	Bad	Good	Neutral	Bad

Please use the box below to give any comments you think are useful for explaining the scores you gave:

**Question 11:**

Do you actually think that [Ebola] will become a global pandemic? I don't know what to think anymore. Is it really rational to think that this will get much worse before it gets better? Or are we slowly actually getting this under control?

Answer 1			Answer 2		
<p>There is no way to know, since this is an emerging pathogen (it never had access to so many people before) and no-one knows where this will end up. It now has plenty of opportunity to optimize transmission in humans. At this point, the best case scenario is that it ends up endemic to most or all of Africa, with sporadic outbreaks globally. But that is dependent on keeping it out of most of the 'third' or 'second' world. Those outbreaks will be small unless we become complacent and stop keeping a close eye on it. After that, it depends on successful development and scaled up production of a vaccine to control or eliminate it, just so long as people comply with vaccination (glaring at you, anti-vaxers).</p>			<p>The worst part of this is that it's going to drag on for quite a while in West Africa, and we're going to get these stragglers with incubating Ebola hit[ting] cities all over the world, over and over again, never knowing when or where they're likely to turn up. That's going to suck but, at this point, it doesn't look as if that's going to lead to any non-West Africa epidemics, thus no pandemics.</p>		
Good	Neutral	Bad	Good	Neutral	Bad
Answer 3			Answer 4		
<p>Well, if you think about it logically for a second and don't panic, you'd realise more people die from AIDS a day than the entire amount of people who have died from this Ebola outbreak. Not to mention malaria, pneumonia and even the flu. This shit will run its course, a vaccine will come round, and this virus will be history.</p>			<p>Looking at the graphs, it appears that the spread of Ebola is an upward curve... that basically means it keeps multiplying. Also looking at the symptoms, it can take up to three weeks for the symptoms to manifest themselves so that could mean it's already spreading in the US. I don't know what the outcome will be. No one does. But it doesn't look pretty.</p>		
Good	Neutral	Bad	Good	Neutral	Bad

Please use the box below to give any comments you think are useful for explaining the scores you gave:

**Question 12:**

If a dog eats vomit from an Ebola patient, can that dog spread it to humans? By now, many will have heard of the Dallas patient having vomited outside his apartment. If a dog ate that vomit, can that dog pass it to humans? If the vomit seeped into the ground, can the vomit still spread?

Answer 1			Answer 2		
<p>Yes, dogs can carry Ebola, this is a concern in Africa where the recently diseased are being eaten by dogs. My first thought when I saw the report of vomiting outside was how common it is for dogs to eat and roll around in vomit. To date, infection from dog to human has not been recorded, infection from animals has come from eating them... however in the US people are in contact with their dog's fluids a lot more than in Africa (we pick up their poo, and let them lick our face, sleep in our beds etc).</p>			<p>Here's [an academic paper], but it's behind a pay-wall...</p> <p>"Dogs and pigs are so far the only domestic animals identified as species that can be infected with EBOV." Weingartl, H.M., Nfon, C. &amp; Kobinger, G. 2013. Review of Ebola Virus Infections in Domestic Animals. In: Roth, J. A., Richt, J. A. &amp; Morozov, I. A. (eds) Developments in Biologicals. Basel, S. KARGER AG, 211–218.</p>		
Good	Neutral	Bad	Good	Neutral	Bad
Answer 3			Answer 4		
<p>And what will panicking help? You are not a little girl so stop acting like one. Act rational.</p>			<p>Dogs are a great indicator species for an area that could be affected by Ebola. They come into contact with many things. They can't easily transmit the virus because they just don't shed it since they're asymptomatic. Additionally, primates are the only animals where Ebola (other than Reston strain) actually causes a symptomatic response in the host. Ebola Reston is completely asymptomatic in humans and does not cause harm.</p>		
Good	Neutral	Bad	Good	Neutral	Bad

Please use the box below to give any comments you think are useful for explaining the scores you gave:

## About you:

---

We would be grateful if you would answer the following questions, but please do not feel obliged to. Please answer only those questions you feel comfortable answering; your survey answers will still be valid even if you leave this section blank.

**Level of medical qualification:** Medical Student | Medical Doctor | Medical Nurse | Medical Scientist | Other [please list]

---

**Age:** 18-24 | 25-34 | 35-44 | 45-54 | 55-65 | Over 65

---

**Gender:** Male | Female

---

**Have you had direct experience of treating Ebola patients?** Yes | No

---

**Where do you think the answers presented here came from?**

---

**Overall, that was your impression of the answers presented here?**

---

**Your contact details\***

---

\* We will only use your contact details to contact you for follow-up relating to this academic study. We will not pass your contact details on to any third parties inside or outside Royal Holloway, University of London



---

**THANK YOU for taking part in this survey.**

**If you have any questions regarding the survey, or are returning it electronically or by post, please send it to:**

Jennifer Cole, Department of Computer Science,  
Royal Holloway University of London,  
Egham, Surrey TW20 0EX;

Email: [Jennifer.Cole.2013@live.rhul.ac.uk](mailto:Jennifer.Cole.2013@live.rhul.ac.uk)

# **Appendix V**

## **Investigation of information quality in r/ebola: Results data**

PHASE II-B: Investigation of information quality in r/ebola.

Results data

(2 pages)

Appendix V presents the data returned by the survey participants in Phase II-B: Investigation of information quality in r/ebola.

**APPENDIX V: PHASE II-B – INVESTIGATION OF INFORMATION QUALITY ON R/EBOLA.  
RESULTS DATA**

Person	Order	QA.a-12.3	QA.b-12.4	QA.c-12.1	QA.d-12.2	QB.a-11.3	QB.b-11.4	QB.c-11.1	QB.d-11.2	QC.a-10.3	QC.b-10.4	QC.c-10.1	QC.d-10.2	QD.a-9.3	QD.b-9.4	QD.c-9.1	QD.d-9.2	QE.a-8.3	QE.b-8.4	QE.c-8.1	QE.d-8.2	QF.a-7.3	
RSM-MM01	F	3	3	3	3	1	3	3	3	3	3	3	3	3	1	1	3	3	1	3	3	3	1
RSM-MM02	F	1	3	2	2	2	1	3	0	1	3	3	2	3	1	3	1	2	3	3	2	2	1
RSM-MM03	F	2	3	2	1	1	1	2	1	1	1	1	3	2	1	2	3	3	3	3	2	2	2
RSM-MM04	F	1	2	1	1	2	3	2	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1
RSM-MM05	F	1	2	1	1	1	2	1	1	1	1	1	1	1	1	2	1	2	2	1	1	1	1
RSM-MM06	F	1	3	1	1	1	1	1	1	1	1	1	2	1	1	2	2	1	2	2	2	2	1
RSM-MM07	F	1	2	2	2	1	2	1	1	1	1	1	1	2	1	2	1	1	2	1	3	1	1
RSM-MM08	F	1	2	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RSM-MM09	F	0	0	0	0	1	2	2	1	1	1	1	1	1	1	2	1	1	2	3	3	1	1
RSM-MM10	F	1	2	2	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RSM-MM11	F	1	2	3	2	2	1	2	3	1	2	2	2	2	1	3	1	0	2	3	2	1	1
RSM-MM12	F	1	2	2	1	2	2	3	3	1	1	1	2	2	1	2	1	2	1	2	3	1	1
RSM-MM13	F	1	3	1	2	1	1	1	2	1	2	1	2	2	1	3	1	2	3	3	3	1	1
RSM-MM14	F	1	3	2	2	1	1	2	2	1	2	2	1	2	1	3	1	2	2	3	2	1	1
RSM-MM15	R	1	2	2	3	1	2	3	1	1	1	2	2	2	1	2	1	2	2	2	2	2	2
RSM-MM16	R	1	2	1	1	1	1	2	1	1	1	2	1	1	1	2	1	2	1	3	2	1	1
RSM-MM17	R	1	2	2	3	2	2	3	2	3	3	2	3	2	1	3	3	2	2	3	3	2	2
Score RSM		19	38	29	29	21	25	31	23	19	24	24	27	27	15	33	22	27	29	36	34	18	
Reddit		B (-3pts)	N (1 pt)	G (+6pts)	G (+3pts)	B (-4pt)	N (1pt)	G (+23)	G (+39)	B (-1pt)	G (+4)	G (+10)	G (+26)	N (1 pt)	B (-3pts)	G (+5 pts)	G (+16)	G (+17)	N (1 pt)	N (1pt)	G (+8 pts)	B (-16pts)	
Score WKCS		10	14	13	14	14	9	16	16	9	16	16	11	21	12	24	15	17	24	30	20	12	
WKCS-001	R	2	1	1	1	2	1	1	3	1	2	2	1	2	1	2	2	2	1	3	1	1	1
WKCS-002	R	1	2	3	2	3	1	2	1	1	1	1	1	3	1	1	2	2	3	3	2	1	1
WKCS-003	R	1	1	2	1	2	1	2	1	1	3	1	1	3	1	2	1	1	3	3	2	1	1
WCKS-004	R	0	0	0	0	0	0	0	0	0	0	0	0	2	1	3	1	2	2	3	2	1	1
WCKS-005	R	1	2	3	2	1	2	3	3	1	2	3	2	1	1	3	1	2	2	3	3	1	1
WCKS-006	R	2	2	0	3	1	1	1	3	2	2	1	2	1	3	3	3	2	3	3	3	2	2
WCKS-007	R	1	2	1	2	2	1	2	2	1	1	3	1	3	1	3	1	2	3	3	1	1	1
WKCS-008	R	1	3	1	2	1	1	3	2	1	2	1	1	2	1	3	1	1	2	3	2	1	1
WCKS-009	R	0	0	0	0	0	0	0	0	0	0	3	1	3	1	3	1	2	2	3	2	2	2
WCKS-010	R	1	1	2	1	2	1	2	1	1	3	1	1	1	1	1	2	1	3	3	2	1	1





# **Appendix VI**

## **Questions for Moderators of [The Reddit xxxx Forum]**

PHASE III: Observation of reddit and interviews with moderators.

Interview guide

(4 pages)

Appendix VI contains the interview guide used to conduct the semi-structured interviews with the moderators of reddit forums during Phase III of the study: Observation of reddit and interviews with moderators.

It is presented on the following pages exactly as it was used during the interviews.



## **APPENDIX VI: PHASE III – OBSERVATION OF REDDIT AND INTERVIEWS WITH MODERATORS. INTERVIEW GUIDE.**

---

### **QUESTIONS FOR MODERATORS OF [THE REDDIT XXXX FORUM]**

Hello, my name is Jennifer Cole. I am conducting this interview as part of my PhD at Royal Holloway, University of London. Thank you very much for taking part – this is of course all entirely optional and you can skip questions/end the interview at any time. Your responses will be anonymised and the data handled with care.

If you have any questions about this research before you feel comfortable to begin, please let me know. If you'd like, please feel free to check out my [university research page](#).

Do you have any questions about this research before we start?

Is it ok if I record this?

*(Semi-structured interview, with questions for face-to-face or telephone/Skype interviews. Interviewees will not be shown the questions prior to the interview, but will be given the option of reviewing the transcript and expanding on/clarifying some of their answers afterwards if required.)*

***Please feel free to answer the questions with as little or as much information as you feel is necessary. You don't have to answer all the questions and can skip any you prefer not to answer.***

### **THEME 1: MOTIVATION AND EXPERIENCES OF MODERATING REDDIT**

*The first part of the interview will cover your experience of Reddit and of being a Reddit moderator in general. Later in the interview we'll move on to questions that are more specific to the [xxxxx] thread.*

- 1. Why do you use Reddit?:**
- 2. How much time, on average, do you spend on Reddit?**
  - a. Each day
  - b. Each week
  - c. Each month
  - d. Have there been times when this has significantly increased or decreased?
    - a. If so, why?

- 3. When did you become a Reddit moderator?**
  - a. Please give approximate date
- 4. Had you moderated other subreddits before this one?**
  - a. Which ones?
  - b. Do you currently moderate any other subreddits?
  - c. Do you currently moderate any discussion groups outside of Reddit?
- 5. Do you still edit other Reddits/forums?**
  - a. If not, why not?
- 6. Had you been a regular poster to Reddit prior to becoming a moderator?**
  - a. To [this forum]
  - b. For how long?
  - c. To another thread on Reddit?
    - a. In a related subject
    - b. In an unrelated subject
    - c. For how long?
- 7. Why did you become a moderator [of this forum]?**
  - a. If you have edited other boards before, why did you become a Reddit moderator of those forums?
- 8. What do you see the main role of the moderator to be?**
- 9. Have you ever had to moderate any factually incorrect or misleading information?**
  - a. What was the information?
  - b. Is the information still online (and if so, can you provide a link to it?)
- 10. If factually incorrect or misleading information is posted to the Reddit thread you moderate, would you expect to become aware of it?**
  - a. How?
  - b. Within what timeframe?
- 11. How do you know if it is incorrect?**
  - a. How do you check or verify its correctness?
- 12. Is it your job to do anything about incorrect or misleading information?**
  - a. What do you do/can you do as moderator?
  - b. How much communication, if any, is needed with other moderators?
  - c. Are you aware of the actions of other moderators?
  - d. Can you see each others' actions?

- 13. What happens, in your experience, to incorrect information within the thread?**
  - a. It is corrected by other users
  - b. It is identified as poor quality by other users
  - c. It is ignored by other users
  - d. It is brought to the moderators' attention
  
- 14. Have there ever been disagreements between moderators regarding incorrect information?**
  - a. If yes, how have they been resolved?
  
- 15. How does the subreddit grow to accept more users?**
  - a. In theory
  - b. From actual experience
  
- 16. What happens if the number of posts becomes too much for you to manage?**
  - a. In theory
  - b. From actual experience
  
- 17. What tools do you use to help in moderating?**
  
- 18. What tools/support would you like?**
  
- 19. Is there any way tools/support for moderators could be improved?**
  
- 20. Do you know the other moderators?**
  - a. In what way?
  - b. Do you communicate outside of moderating this subreddit?
  - c. Do you moderate other subreddits together?
  
- 21. Is there a "chief moderator"?**
  - a. If yes how did s/he get to become chief?
  - b. Are some moderators more senior or authoritative than others?
  - c. If so, how did they get there?
  
- 22. What do you think about medical information on Reddit? Do you think Reddit is a good place for medical information?**

## THEME 2: POTENTIAL USE OF REDDIT IN PHEICS

*Finally, we are interested in your views on whether you think Reddit, or a discussion forum like Reddit, could be a useful way of enabling people to find health information during a serious disease outbreak where the usual access to professional healthcare is stretched – such as if, in an unlikely scenario, the Ebola outbreak had spread significantly to the US or UK.*

*In such a theoretical scenario, we are interested in how you think people might use Reddit, and how this might affect your role as moderator. Information people might seek during such an outbreak may not only be about the disease responsible for the outbreak; it may be that they are finding it difficult to reach a doctor for other health issues while the medical staff are overworked dealing with the outbreak. For instance, during Ebola, deaths from malaria increased significantly; midwives were not always willing to attend births; people were unwilling to attend hospitals for physical injuries or sickness if they thought it might bring them into contact with Ebola patients, etc.*

*Reddit might be irrelevant and useless in such a scenario or it might be of some use. What do you think?*

*We are considering which of Reddit's characteristics would make it rather useful or quite useless in such a scenario. What do you think?*

- 23. Do you think Reddit would be able to help people during a serious disease outbreak or not? (If so, how?, if not, why not?)**
- 24. What part, if any, do you think Reddit might play in a serious disease outbreak in a country such as the US or UK?**
- 25. Is there anything else you would like to tell me that you think may be useful or interesting to my study?**

Would you be happy to give me your email address so that I could contact you again for further information?

Would you be happy for me to follow up with another interview if there was anything I wanted to clarify? **YES/NO**

Would you like me to email you a summary of my findings? **YES/NO**

*Thank you very much! If you would like any further information, please feel free to contact me at [Jennifer.cole.2013@live.rhul.ac.uk](mailto:Jennifer.cole.2013@live.rhul.ac.uk)*



