1 The Internet, Warts and All

> Invention, it must be humbly admitted, does not consist in creating out
> of void, but out of chaos . . .
>
> Mary Wollstonecraft Shelley, Introduction to *Frankenstein*

1.1 Warts and All

When Oliver Cromwell went to Samuel Cooper to ask for his portrait to
be painted, he was the most powerful man in the country: Lord Protector
of England. He had effectively deposed King Charles I: the king was
executed in 1649, the same year that Cromwell invaded Ireland and
perpetrated the massacres that make him one of the worst of all villains
in the eyes of the Irish. He was a driven man, a 'great' man, but still a man
almost obsessed with humility and what he saw as the truth. As John
Morrill put it: 'He was called to overthrow tyranny and pride and replace
it with humility and a common concern to share the fragments of truth
that so many men of goodwill had been granted.'[1] This obsession with
humility and with truth presumably lay behind his answer to Cooper's
question about whether to include his prominent and disfiguring warts in
his portrait. It would have been easy to have the portrait exclude them –
indeed, the most prominent portraits of Cromwell, from an initial portrait
by Cooper himself in earlier years to the later much-copied full-sized
portrait by Sir Peter Lely, seemingly based on Cooper's 'warts and all'
miniature, have the warts blurred, minimised or removed.

For the crucial Cooper miniature Cromwell was clear. He wanted the
portrait to show him 'warts and all'. He wanted as truthful a picture as
possible. He wanted to be remembered as he was, not in some false,
idealised form. This is how the term 'warts and all' entered the English
language – to see a thing as a whole, including all the unappealing
features. If we are to understand something properly and truthfully, we

[1] Morrill 2007, p. 121.

1

need to see, understand and accept its unappealing features as well as those features that we find attractive.

This story about Samuel Cooper, recounted above as though it were fact, may actually not be true. Some attribute it to another artist entirely – the aforementioned Sir Peter Lely – based in part on Horace Walpole's famous book of anecdotes collected more than 100 years later.[2] The whole story may be apocryphal. It may be 'fake news' invented after the events in order to portray Cromwell in a favourable light. It is all but impossible, more than 350 years after the event, to be absolutely sure what actually happened or what was said. Expert opinion based on other historical evidence and a detailed analysis of the paintings themselves strongly suggests that it was Cooper to whom the 'warts and all' comment was made, but it is not absolutely certain.[3] That is generally the nature not just of history but of much more. Certainty is rare. That is equally true of much concerning the internet.

1.1.1   The Internet We Have

If we are to get closer to the kind of internet we want, we need a better understanding of the internet that we have. We have to look at the Internet, warts and all, and not be seduced by the illusions of how the internet *seems* to be, or how others would like us to see the internet for their own purposes.

The internet is not a permanent, perfect archive of historical information, indexed by neutral and purely algorithmic services such as Google or accessed through neutral and apolitical platforms such as Facebook. It is messy, dynamic and constantly changing. Things are being deleted, modified and blocked all the time. Search engines, social media platforms and other services are not neutral public services but money-making self-serving businesses. Their algorithms are neither neutral nor 'organic', but created by humans and with biases, assumptions and faults. Intervention with those services and platforms, and with the algorithms and indexes

---

[2] In Horace Walpole's *Anecdotes of Painting in England, with Some Account of the Principal Artists*, p. 226 of the 6th edition, Walpole records an indirect anecdote: 'Captain Winde told Sheffield, Duke of Buckingham, that Oliver certainly sat to him, and, while sitting, said to him, "Mr Lely, I desire you would use all your skill to paint my picture truly like me, and not flatter me at all; but remark all these roughnesses, pimples, warts, and everything as you see me, otherwise I will never pay a farthing for it".'

[3] See, for example, the catalogue of Phillip Mould's exhibition, 'The Portrait Miniatures of Samuel Cooper (1607/8–1672)'. The Cromwell portraits by both Cooper and Lely are Cat. 21–23, with detailed commentary by art historian Dr Bendor Grosvenor, pp. 70–74: http://philipmould.com/application/files/3114/4708/8432/Warts_and_All_catalogue_v12 .pdf.

created by them, is not a fundamental and wholly inappropriate inter-
ference with freedom of expression, but part of a regular, important and
potentially positive process that can help keep a more appropriate balance
between the rights and interests of people and corporations.

The internet is neither an ungoverned and ungovernable realm of
criminals and terrorists that needs to be reined in to protect us nor
a massive surveillance engine that has brought us to the brink of an
Orwellian dystopia. There are criminals and terrorists – and paedophiles
and drug dealers – on the internet, but most of the time, for most of the
people, it is a place to find information, socialise, do business, and gen-
erally live, and do so in relative safeness and simplicity. There is a great
deal of surveillance – most of people's web activity is monitored in a wide
variety of ways – but the surveillance is neither as effective nor as mal-
icious as some might suggest. Understanding the context, the complexity,
the nuance, the dynamism and the relationships between the various
issues – and, in particular, understanding the messiness of the whole
situation – can help us to take a more balanced view of each of the issues
in turn.

1.1.2   Free Speech, Privacy and Truth

Free speech, privacy and truth are the central themes of this book.
There are specific chapters devoted to each of them – Chapter 5 on
Free Speech, Chapter 6 on Privacy and Chapter 9 on Truth – but
none of these is an idea about which it is easy to be precise. Neither
free speech nor privacy can be easily defined or pinned down. Some
scholars contend that attempting to define privacy in particular can
be counterproductive or a distraction from addressing the very real
problems.[4] Whatever definition is taken can end up either missing
something crucial or covering areas that are really not about what
people understand by privacy at all.[5] Pinning down free speech may
be just as difficult. What counts as 'speech' and what constitutes
'freedom' is not as simple as it seems. Freedom to do what?
Freedom from what? Truth may look as though it is easier to under-
stand and deal with, but even here there are difficulties. Perspectives
matter. Interpretations matter. Context matters. All these things are

---

[4] For example, Helen Nissenbaum: '[b]elieving that one must define or provide an account
   of privacy before one can systematically address critical challenges can thwart further
   progress'. Nissenbaum 2010, p. 2.
[5] Daniel Solove, notes that '[p]rivacy is too complex a concept to be reduced to a singular
   essence. It is a plurality of different things that do not share one element in common but
   that nevertheless bear a resemblance to each other'. Solove 2011, p. 24.

discussed in more depth in the relevant chapters, both in general and in the specific topical examples that are examined – social network-ing, trolling, fake news, surveillance and so forth. It is a central contention of this book that the three central issues of free speech, privacy and truth are intrinsically and inextricably linked, in a way that is often surprisingly complex and nuanced.

1.1.3   Taking a Broad Look

This is a book about law. It is also a book about technology, about politics, about psychology, about society, about history, about philosophy. A great deal is covered here, by design. We need to take a broader, more multi-faceted approach to the way we deal with the internet. This means, amongst other things, that we need to consider all these aspects. Decisions in relation to the internet that may *seem* to be about law have political implications, technological implications, societal implications and more.

Legal and technological measures that impact upon one of the three key issues – free speech, privacy and truth – will generally have an impact upon another – or more likely on *all* – of them, and often not in the way that appears immediately obvious. For example, a policy such as requiring real names on social media,[6] whilst osten-sibly about authenticity – and hence truth, will mean invasions of privacy and will chill freedom of speech for many. Sometimes it will reduce truthfulness as people unable to gain the privacy-related pro-tections of anonymity or pseudonymity will be more likely to omit some of the truth or to lie. As the many examples presented in this book will demonstrate, 'real names' is not the exception but the rule: free speech, privacy and truth cannot be easily separated.

Other examples show the same patterns. Invasions of privacy are used to identify websites to censors. Trolls 'dox' people, invading their privacy and revealing true details of their lives in order to scare them into silence. The advertising industry argues that ad-blockers – installed by some specifically to protect themselves against the inva-sions of privacy that are inherent in behavioural targeting systems now commonly used by advertisers – are an infringement on their freedom of speech.[7] The same pattern is repeated again and again. Whichever of the issues we are trying to 'deal with', we need to consider each of the others.

[6] See Chapter 8, pp. 220–223.   [7] See Chapter 9, p. 257.

1.2   Perspectives on the Internet

We are confused about what the internet is, but we are also confused about what we want the internet to be. The two confusions interact to produce even more confusion: it is that interaction that makes the regulation of the internet especially difficult and particularly troublesome.

1.2.1   The Internet as an Information Resource

The ability to find information is a critical but often misunderstood and underestimated part of freedom of expression: in most formulations, the right to freedom of expression includes the freedom to both impart *and receive* information.[8] The internet can help this freedom to be realised perhaps better than anything else in history – both in theory and in practice. If you want to find something out, the first thing that you do is look on the internet. Almost any kind of information can be found – anything that can be digitised, from the text that formed the early web to images, sound, video, 3D imagery and much more. You are most likely to *search* for the information – and most often to use Google for that search.[9] You may instead go through a social media platform – more than two billion people are on Facebook.[10] You might use an online encyclopaedia – most likely Wikipedia, which has more than 43 million pages of regularly checked and updated information on almost everything.[11] You might look at specific and specialised websites for particular subjects, at forums that you know about, or ask people that you know (either online or in 'real' life) and trust to recommend places to visit.

There are conflicting needs when using the internet as an information resource. For the purposes of historical research, an archive should be keep as pristine as possible, with records as complete as possible and as they were when they were laid down. Records should not be deleted, and any modifications made should be recorded, together with the reasons for them – and only certain kinds of modifications should be possible: adding newly discovered resources, for example, or correcting clear errors. History, in these terms, should not be 'rewritten'. In principle, this sounds nice and clear, but even a little thought shows that it is not as simple as that. What kinds of errors should be corrected? Who should determine

---

[8] E.g. in Article 10 of the European Convention on Human Rights and in Article 19 of the Universal Declaration of Human Rights. See Chapter 5, p. 104.

[9] Google has dominated the search market for some years – around 90 per cent worldwide. See e.g. http://gs.statcounter.com/search-engine-market-share.

[10] Mark Zuckerberg celebrated Facebook passing the two billion people mark on 27th June 2017. See www.facebook.com/zuck/posts/10103831654565331.

[11] See https://en.wikipedia.org/wiki/Wikipedia:Size_of_Wikipedia.

whether an error has been made and how to correct for it? If an error is
discovered, how can the repercussions of that error be addressed as well as
the error itself? If it is determined that it was definitely Samuel Cooper
rather than Peter Lely to whom Cromwell gave the instruction to paint
him warts and all, how can all those records (dating back to 1764 at least)
that suggest it was Lely be corrected? If they are not corrected, people will
continue to be misled. They will not find the truth. If they are corrected,
changes will have to be made in the historical archive itself. Should
Horace Walpole's seminal piece of work on English painting be marked
down as including an error everywhere it is mentioned?

   There are no easy answers here, primarily because the idea of an
accurate and complete archive is based on a misunderstanding of the
nature of history and the nature of factual information. Not only do new
facts emerge but our understanding of existing facts and interpretation of
them changes. As J. S. Mill put it, '[v]ery few facts are able to tell their own
story, without comments to bring out their meaning'.[12]

   All this means is that a historical archive – or, to be more precise, an
information resource useful for historical research – can often be com-
plex, with notes, qualifications, references and cross-references. It can
need time, attention and expertise to understand and to navigate
through – but the complexity is necessary for the archive to be useful.
That is fine for those with the requisite time and expertise, but it makes
the archive very much less useful for those without it – and the vast
majority of users of the internet have neither the time nor the expertise.
They need a very different kind of information resource: they want
information quickly and easily, in a form that can be understood without
specialist knowledge or expertise.

   For these people – most people – if the required information exists but
is hard to find, or if would take too long to find, or if it is obscured by
a morass of other information, or if verifying that it is the right information
is too difficult, then for most practical purposes it might as well not exist.
People trying to find the information without sufficient time, energy and
expertise will not find it. This is one of the reasons that internet tools like
search engines have become so popular: they make it easy and fast to find
things for everyone. People want to find the most relevant information
quickly, and don't want to be bothered with detailed fact checking – or
even know how to do proper fact- or source-checking, one of the reasons
behind the fake news problems.[13] They aren't likely to want to have to go
through information in detail before finding the pertinent facts – they may

---

[12] Mill 1859, p. 22.   [13] See Chapter 9.

well only look at headlines on search results without even visiting the relevant pages, for example.

People in this scenario want their search engines, social media platforms or online encyclopaedias to help them to find the information they need. They want those engines, platforms and encyclopaedias to be *neutral*. At this point they do not seem to be aware that these engines, platforms and encyclopaedias are not, and *cannot* be, neutral: the crucial neutrality myth is the subject of Chapter 4. They don't want to be distracted by irrelevant information. They don't want old or irrelevant information to clutter up their timelines or search results: having complete and accurate information, as required for historical archives, could be a distraction and a disadvantage. It could stop them finding what they want to find.

The search engines and social media platforms know this. This is why Google constantly tweaks its algorithm and has been 'personalising' search results based on what it 'knows' about individuals since 2009,[14] and why Facebook has been algorithmically curating its news feed since it was introduced in 2006.[15] Twitter, whose nature until that time had been much more about a pure, unaltered, chronological timeline, began experimenting with algorithmic curation of timelines in 2015, and has tried a number of different versions, including 'while you're away' (a curated selection of tweets since you last logged on) in January 2015,[16] 'Moments' (thematically curated tweets) in October 2015[17] and 'Best Tweets' (tweets you're most likely to care about) in February 2016.[18] These have been as options rather than the default timeline, but they emphasise that the need to select or curate, to help people to have a route through the mass of information to what they need, is seen as vital for the mass audience. The needs of most people for speed, convenience and ease of use are quite different from the needs of specialists – journalists, academics, researchers – for a complete, accurate and historical archive that can be trawled through exhaustively and logically. When both groups want access to the same information and use the same tools – search engines, social media platforms and so forth – that naturally brings tension, confusion and problems.

[14] https://googleblog.blogspot.co.uk/2009/12/personalized-search-for-everyone.html.
[15] See the official launch notification of Facebook's News Feed and Mini Feed: www.facebook.com/notes/facebook/facebook-gets-a-facelift/2207967130.
[16] https://blog.twitter.com/official/en_us/a/2015/while-you-were-away-0.html.
[17] https://blog.twitter.com/official/en_us/a/2015/moments-the-best-of-twitter-in-an-instant-0.html.
[18] https://blog.twitter.com/official/en_us/a/2016/never-miss-important-tweets-from-people-you-follow.html.

When looked at from the perspective of those wanting 'their' informa-tion to be known rather than those seeking information, a whole set of other potential issues arises. First of all, which people are concerned – those who create or want to disseminate information, or those who the informa-tion is about? Others who might be impacted upon if the information is known? Still more who just have an interest in a subject or an agenda? Some people will want particular information to be found by everyone. Some would prefer this information not to be found at all. Others would like it to be found by some and not by others. When information concerns more than one person, their desires may be in tension or in conflict.

The internet's role as an information resource also brings in the need for privacy. For an information resource to function well it not only has to exist and be relatively 'user-friendly', it has to actually be used – and that means that people need to be willing to use it. They need to know that their use of the information resource will not in itself be used against them. A victim of spousal abuse will not search for information about refuges if they believe their abuser could discover they were searching for them and even discover which refuges they have been investigating. A whistle-blower would be wary of putting information about their employer's misdeeds on the internet if they thought their employer might easily be able to discover who they are. A teenager might not seek out information about sexual health if they thought their conservative parents would immediately know it. A dissident would not want their oppressive government to know that they were accessing opposition websites or information that showed that government in a bad light. One of the most notable results of intrusive surveillance is a drive towards conformity and an unwillingness of people to take risks.[19]

There are others whose interests come into play here: groups, most importantly governments and other authorities, who wish to prevent people having access to information for various reasons. Access to offen-sive material such as child abuse imagery,[20] material deemed to promote terrorism or encourage extremism, material which breaches copyright, material that should only be accessed by adults, or material that is deemed defamatory, blasphemous or offensive to others. The variety of different reasons that material might be deemed objectionable by one government or another, or by one interest group or another, is extensive.[21] Sometimes

[19] See Chapter 7, particularly pp.135–136.
[20] Historically (and in places such as the USA) referred to as 'child pornography' – the term 'child abuse imagery' is generally preferred in the UK as it makes clear that even the making of this material involves child abuse.
[21] Discussed in Chapter 5, and in some specific areas in Chapters 8 (on trolling) and 9 (on fake news).

the issues and reasons behind the blocking are political, sometimes moral or religious, sometime economic, sometimes pragmatic or instrumental. Some of the reasons are positive and valid – most formulations of freedom of expression include caveats such as for the protection of people or of their rights – but some are distinctly less convincing. The balancing of rights and interests in conflict is often complex and nuanced.

At the moment, maintaining that balance is largely in the hands of the internet giants who control so much of the access to the information – Facebook, Google, Twitter and, in a rather different way, Wikipedia. In whose interests do these organisations operate? That of their 'customers'? Of their advertisers? Of their shareholders? Facebook, Google and Twitter are businesses and the bottom line is the bottom line, which sometimes means that people's rights and needs do not exactly take centre stage.

How regulators could or should respond to that is complex. Lawmakers can tend to take the rights of the third group – those who wish to block access to material of various kinds – more seriously than others, primarily because they themselves are often in that position, and – as shall be demonstrated by many of the examples in this book – because they have a limited understanding of both the issues and the technology. It is important to be fair to the lawmakers, however: this is not easy. The balances are very difficult to find even when the issues and technology are understood, but it is of critical importance and could often be done much better.

1.2.2   The Internet as a Communications Medium

At its beginning, the internet was primarily a communications medium, and that aspect has remained and dramatically expanded over the years. Electronic mail (email) was one of the first applications for the internet and remains one of the most important and trusted.

Some communication is effectively instant and ephemeral: Internet Relay Chat ('IRC'), the first system in common use, was developed in 1988.[22] Others, from email onwards, are intended to form part of a permanent record. Official and legal correspondence is often done by email – and is expected to be part of official records. Emails are subject to freedom of information law:[23] there is a good reason why Hillary Clinton got into so much trouble for seeming to hide and then delete a significant

---

[22] A summary of the history of IRC is online at https://daniel.haxx.se/irchistory.html.
[23] In the UK under the Freedom of Information Act 2000. There are similar laws in many states.

amount of email correspondence. In the current internet, a vast variety of
forms of communication are possible, from the equivalent of telephone
calls (Voice over Internet Protocol – 'VoIP' – has been in relatively
common use since 2003) and video calls (Skype launched its video calling
system for Windows in 2006)[24] to experimentation in 3D virtual reality
communications.[25]

The requirements of the internet as a communications medium are
qualitatively different from that as an information resource. Someone
communicating directly with another person needs to know that their
information has gone safely and securely to the right place, fast enough
and reliably enough for the particular kind of message. Instant and inter-
active communication puts the emphasis on speed – bandwidth was the
key limitation for early adopters of online video communication – and
reliability of connection.

Communication also brings privacy into play. Different kinds of com-
munication require different kinds of privacy. Some are highly confiden-
tial – the use of encryption for communications has a history far older than
the internet – whilst others might be readily shared within various differ-
ent groups who hold the requisite trust. Privacy in law has generally
considered privacy of correspondence a key element. It is included in
both the Universal Declaration of Human Rights[26] and the European
Convention on Human Rights,[27] and the US Supreme Court ruled it was
constitutionally protected as early as 1877.[28] Opening letters, tapping
phone lines and their equivalents are not things that can be done as
a matter of course in a democratic state.

1.2.3   The Internet as a Business Platform

Though the internet was initially a communications platform for the
military, scholars and geeks,[29] the opportunities that it presented for
business became apparent relatively quickly. The rapid growth and devel-
opment of the internet over more recent years could be argued to have
taken place to a great extent because businesses have grasped those
opportunities. There is a reason why many of the biggest companies in
the world are primarily internet-based companies. In 2017, according to
Forbes, the top four companies in the world in terms of market

[24] See https://blogs.skype.com/wp-content/uploads/2012/08/skype-timeline-v5-2.pdf.
[25] E.g. Facebook Spaces; see www.oculus.com/experiences/rift/1036793313023466.
[26] Article 12.   [27] Article 8.
[28] In *Ex parte Jackson* 96 US 727 (1878), online at https://supreme.justia.com/cases/federal/
us/96/727/case.html.
[29] See e.g. Naughton 2000.