THREE

Digital Safety Debates and Regulatory Environments

Introduction

The Online Safety Act 2023 (OSA) is a landmark piece of legislation that attempts to establish a new regulatory framework to make the UK 'one of the safest places to be on the internet' (Department for Digital, Culture, Media & Sport and Home Office, 2019). The OSA establishes a new 'duty of care' for digital service providers towards their users, obligating them to act against harmful content published on their sites/platforms and to provide appropriate protections for users, based on vulnerabilities. It is unsurprising that the OSA's implementation was deemed controversial, especially by technology companies, given its wide-ranging regulatory requirements for digital service providers, such as social media platforms and other major technology companies. The significant regulatory change brought about by the OSA is captured by Nash and Felton who rightly contend that the Act has 'tried to embrace an ambitious new model of regulation that could rebalance corporate conduct amongst the biggest technology platforms in favour of the public interest' (2024, p 13).

Given the change required, it is unsurprising that tensions and contentions surfaced during the OSA's pre-legislative development phases arising from the publication of an initial White Paper (Department for Digital, Culture, Media & Sport and Home Office, 2019) that formed the basis for both a draft

Online Harms Bill in 2021 and then a final Online Safety Bill in 2022. Many of these tensions involved the issue of who will take responsibility for implementing the required regulatory measures, including efforts by key industry stakeholders to shape future legislative application in line with perceived interests and business models. The primary areas of contention centred around issues of content moderation, platform design (and algorithmic accountability) and the individual/collective harms that are created by these platforms and their content, as well as their effects upon society at large.

Many of these debates were captured in the more than twenty Joint Committee hearings where several key stakeholders (politicians, regulators, rights campaigners, researchers and technology companies) discussed these issues in-depth and allowed for the scope and expected impact of the Bill to be tested through expert witness testimonies. From these exchanges, the debates around the Online Safety Bill, and online harms in general, are interwoven with discussions around duties, obligations and the accountabilities of certain individuals/stakeholders.

This chapter examines the fractious debates entailed in the Bill's passing. Core tensions are identified through a thematic analysis of the Bill's hearings published in Hansard. Two specific areas of contentions are investigated: (1) age verification and (2) scam advertising. Examination of key testimony in these two contentious areas highlights different attempts by the state and industry to allocate and avoid responsibility. The analysis points to how digital responsibility is established and realised at state-level. It identifies how digital responsibility manifests within the different dimensions of digital product design, delivery and use by the private sector, including different responsibility contentions that underpin the debates. Doing so places a spotlight on the roles that the core technologies of authentication and access control play in forcing the actioning of digital responsibilities by both the end users of digital technologies and by technology providers.

The chapter finds that a responsibility consensus across key parties was not fully realised which meant that legal responsibility was established through the implementation of the OSA. Legislation was used to essentially command that responsibility was discharged through technological means. The Parliamentary debates thus demonstrate diverse types of responsibility action at play, which were mostly negative, particularly from the private sector parties, in trying to avoid responsibility for online harms and safety.

Background

The broader policy intention of the Online Safety Bill was to create an international benchmark and gold standard for online regulation of harmful content, at a time when it was becoming clear that harms were increasing and self-regulation was not working (Coe, 2022). The Bill was praised by some MPs for tackling concerns about the internet, increasingly perceived akin to a 'Wild West online' (BBC, 2021). The lack of content regulation had left users, and vulnerable persons in particular, open to abuse, violence and fraud. At the same time, digital service providers were seen as benefiting from the perceived Wild West, as it allowed them to operate without impunity and develop business models that placed profit over safety. That said, the Bill was not without its detractors, with some digital rights campaigners outlining concerns over the overarching governmental powers that could give rise to state-backed censorship (Burns, 2021a; Smith, 2020).

It is in this rather febrile atmosphere that the OSA began its legislative passage, which itself was protracted (McGlynn et al, 2024). Prior to the OSA, the UK Government held two rounds of public consultation in 2017 and 2019 which led to the development of the initial Online Harms Bill, and then the subsequent Online Safety Bill, in 2022, which eventually led to the OSA (Nash and Felton, 2024). The change in focus from harms to safety is significant as it belies a change in government policy about who should be responsible for the prevention of online harms and how it should be done (Nash, 2019). The change represents a shift from *ex post* allocation of responsibility based on harmful content to an *ex-ante* approach of safety. The new approach goes beyond solely harm-based content to consider the potential impacts of business models more broadly. The systems-based perspective at the heart of the OSA gives rise to significant responsibilities for digital service providers and allows the state great scrutiny powers to examine and shape business model development (Nash and Felton, 2024). Moreover, the OSA's development also encountered troubled political waters that involved a backbench revolt and significant amendments proposed by the House of Lords which eventually shaped the OSA's final construction (Nash and Felton, 2024).

Unsurprisingly, the nature of harm underpinning online harms was an area of contention. The Online Harms White Paper of 2019 covered a panoply of harmful online situations ranging from national security to individual harms. The breadth of potentially harmful situations meant that the basis of harm relevant to government actions was unclear. Imposing a duty of care relating to harmful situations online based on content, when the harm is manifest and unclear, was challenging (Woods, 2019). A different regulatory approach was required and the 'starting point' of harmful content was wrong (Perrin and Woods, 2018). Regulatory action should instead focus on the broader online environment created by service providers and what providers could do to make that environment safer. The environments created by online service providers were viewed as public spaces and needed to be regulated with the same expected duties (Leiser and Harbinja, 2020). The greater focus on the online environment through a systems-based approach was beneficial because it better aligned responsibilities (Nash and Felton, 2024). This type of public space governance means the online spaces provided by service providers are not seen as technologically neutral. Instead, providers can use these spaces to shape user activities and actions, including for the profit motives of the provider.

The increased regulatory focus on environments, including business models, could better outline the risks arising from the design of digital services. It would also better align provider responsibilities with potential individual harms to ensure providers were responsible for ameliorating risks, especially to vulnerable users. Online providers would therefore be made more accountable for the potentially harmful content posted on their services (McGlynn et al, 2024) which would lead to a more positive digital experience for everyone (Nash and Felton, 2024). Greater accountability would create clearer processes of control relating to flows of harmful content. Online service providers would therefore be responsible for the spaces that they govern and manage the hazards generated within their space, including from how business models are constructed (Price, 2021). Doing so places responsibility on online service providers to implement greater protective measures as part of their core business models and operations.

The new regulatory focus was based on health and safety regulation which had a much greater focus on business model design and the core idea that providers have a moral responsibility to consider potential harms that may arise to users and to actively mitigate against them (Bearitz and Schertel Mendes, 2023). The imposition of a statutory duty of care thus involved proactive management of defined risks arising from providers rather than a universal duty to not do harm (Woods, 2019). The regulatory duties required by different service providers were thus differentiated and reflected the risks arising from individual services and their degrees of scale that could extenuate harms for users (Tambini, 2019).

The OSA identifies distinct types of service provider. Category 1 providers that entail high risk, high impact for users, such as pornographic services, have the highest obligations. Category 2 providers are high reach services, such as major search engines, or providers that entail higher

Figure 3.1: A new regulatory approach

HEALTH & SAFETY
The internet needs to be a SAFE SPACE!
WHOSE responsibility is it to clean it up & make it SAFE?

risk in their function or operation, because a certain type of data is a core part of their business model. Accordingly, the duty of care imposed is differentiated because the first duty is based on provider and the second is based on content (Tambini, 2019). The types of legal responsibility that arise thus depend on the type of provider, their scale and content that is generated on their platform. That said, all providers have duties relating to the prevention and removal of illegal content and/or content that could be harmful to children. The child safety requirement gave rise to a greater focus on age verification which became one of the key contentions outlined in the bill passage debates.

The Bill thus established a duty to protect in different ways and for different types of provider audience. Three duties were initially proposed (Trengove et al, 2022) to: (1) protect all users from illegal content, (2) provide additional protective measures to keep children safe, if the service was likely to be used by children, and (3) protect all users from content that is harmful without being illegal, if the service was of sufficient reach and magnitude. The latter proved particularly controversial and was not included in the final draft of the OSA. Nevertheless, the OSA, as implemented, features a significant degree of the Bill's components.

The OSA regulates providers by establishing systems-based responsibilities to ensure positive digital experiences, especially for children. Service providers, under these duties, 'bear a preventative duty of care to users in the design and delivery of online services' (Nash, 2019). The OSA imposes duties of care on both types of provider in relation to two broad types of duties: (1) to conduct risk assessments based on potentially harmful content and (2) to mitigate potential risks arising. Consequently, even though the OSA takes a broader environmental approach, it nonetheless still retains aspects of its initial focus on harmful content, except now as part of the broader imposition of duty-based statutory responsibilities. Risk assessment is thus a key part of the responsibilities expected. Risk assessments regarding illegal content are required by the OSA for service providers to factor in freedom of expression and privacy. Services provided for children have additional responsibilities for risk assessments, to protect users from harmful content (Nash, 2019). The OSA also has extra jurisdictional application that covers service providers with links to the UK, that has UK users as part of a target market and which can include a service with a significant number of users or a service available in the UK that may give rise to a material risk of significant harm to individuals in the UK (McGlynn et al, 2024).

Taking on board these points, the OSA requires regulatory input which is primarily provided by Ofcom. Ofcom is tasked under the OSA to provide guidance on risk assessments and devise codes of practice in relation to the application of duties. Compliance with Ofcom derived codes fulfils the OSA safety duties for regulated service providers (McGlynn et al, 2024).

Ofcom can also monitor compliance with codes of practice and can enforce alternative measures to better meet Ofcom guidance, if a service provider is deemed to not be meeting expectations. Ofcom's strongest powers regard investigation and enforcement where it can issue corrective orders and impose fines if a code breach is demonstrated. Ofcom can fine companies up to £18 million, or 10% of annual global turnover, whichever is higher (Coe, 2022). In that sense, even though the OSA is in effect a legislative command, it is nonetheless representative of a hybrid regulatory approach which blends elements of public and private governance (Bearitz and Schertel Mendes, 2023).

In summary, the OSA's focus on online service providers' design and operation reflects a greater focus on business model construction and application as a key cause and generator of harms. The Online Safety Bill gave rise to a new systems-based approach to online regulation of online harms that does not focus on the removal of a particular type of content per se. Regulatory response and compliance activity is more about a provider's underpinning infrastructure and operations that gives rise to systemic issues which could have greater impact through amplification of content, especially through advertising-based business models. The key consideration that links to business model application is a direct recognition that certain types of business action are designed to influence user behaviour. Online service provider business models and their operation are not technologically neutral. The OSA pre-legislative journey recognised that they were designed to achieve a desired outcome and that outcome can have a negative and harmful impact on users and the broader societal environment. As such, the nature of harm that underpins the OSA's application shifted from one that focused on individual harms to one that recognised the societal responsibilities that online providers should have to create and develop online spaces safely (Trengove et al, 2022). However, the shift to a broader and more abstract notion of harm generation, with its focus on environmental space and community dimension, means that the ascription of individual responsibility, including to service providers, is potentially more challenging, as outlined in the key debates leading up to the Bill's passing.

Research approach

To identify the responsibility contentions that arose from fractious pre-legislative debates, we examined the four key stages and legislative components of the Bill. They were:

1. The Online Harms White Paper – published by the UK Government in April 2019, the White Paper sets out the government's initial proposals for the Bill (Department for Digital, Culture, Media & Sport and Home Office, 2019). Resulting from commitments made in the Internet Safety Strategy Green Paper (Department for Digital, Culture, Media & Sport, 2017), the White Paper also includes a public consultation.
2. The draft Online Safety Bill – published by the UK Government in May 2021.
3. We also draw on expert oral evidence provided during the Joint Committee on the Draft Online Safety Bill between September and November 2021 (referenced in the paper as 'JC#').
4. The Online Safety Bill itself, which was published by the UK Government in March 2022.

Documents were coded thematically to identify underlying themes and issues relevant to responsibility actions and concomitant contentions. Our initial analysis of relevant documents highlighted two issues of contention arising from pre-legislative debates: age verification requirements and scam advertising.

We focused on these two debates in greater depth as they highlighted significant discussion and contention about the need for online service provider actions and responsibility for both concerns and ameliorative processes. Equally important, both concerns highlighted debate about how digital responsibility can manifest in private sector product design. We outline both concerns in more depth.

Debate#1 – age verification

One of the 'landmark' proposals from the Bill aimed to tackle underage access to pornography. Relied-upon research demonstrated that approximately 51% of children aged 11–13 had seen pornography (Department for Digital, Culture, Media & Sport, 2022). Children's charities also expressed concerns about the damaging impact pornography can have on children's health, behaviour, relationships and the risks associated with child abuse and exploitation (Barnardo's, 2021). The Bill mandated all pornography websites, as well as platforms that publish and place pornographic material on their sites, to introduce age verification technologies and processes to access for users under the age of 18, replicating the policy direction relating to online gambling. The proposal was also in line with protections outlined in the Digital Economy Act 2017 and the Information Commissioner's Office (ICO) Age Appropriate Design Code (2020) which is a code of practice for online platforms providing services relevant to children. Verification checks would not require a full identity check and, by ensuring such technologies are 'secure, effective and privacy-preserving' (Department for Digital, Culture, Media & Sport, 2022), they should not process/store any data that is not relevant for the purpose of age verification.

The Bill's age verification proposals were largely welcomed, particularly by children's rights groups, but they also sparked ancillary debates over the scope of the legislation and the powers it affords. For some, the proposal threatened 'the integrity of the Internet's architecture' (Burns, 2021b) while others contended that the Bill's scope did not go far enough.

For example, children's charity Barnardo's warned that the Bill's categorisation of online service providers could also 'lead to a two-tier system that means children [could be] potentially able to access harmful content through these smaller sites' (Barnardo's, 2021).

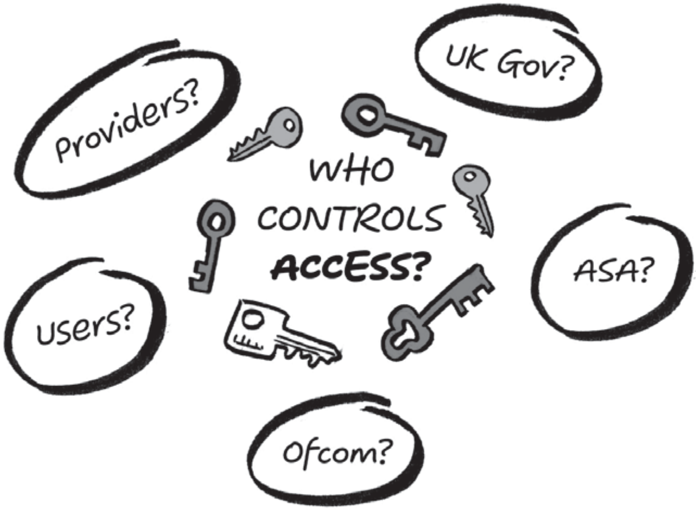
The new focus on age verification, and the technological controls required to implement this, raised issues about the distribution of responsibility that led to fractious debate between private sector providers, government and civil society. The proposal was perceived as a shift of responsibility from the state to online service providers to provide age verification services as part of their core business model. The proposed new allocation of age verification responsibilities onto a new, emerging market was one that had been sought by civil society activists over previous years (Burns, 2021b). While this new form of responsibility allocation was set in train by the Bill, it was not absolute because 'the onus will be on the companies themselves to decide how to comply with their new legal duty. The bill … is flexible to allow for innovation and the development and use of more effective technology in the future' (Department for Digital, Culture, Media & Sport, 2022). Online service providers were therefore given leeway to select technological solutions appropriate to their circumstances including the impact on their overall business model. The delegation of technical solution to online service providers was criticised as it would allow providers to choose age verification controls that are more favourable to their business model and approach. The allocation of age verification responsibility and the degree it should be actioned by providers was a key element of this fractious debate.

Debate#2 – scam advertising

The Bill also proposed new solutions to tackle scam ads or online ad fraud that had become 'endemic' over the previous decade (Cookson, 2016). Ad fraud and scam ads grew dramatically during the COVID-19 pandemic, with digital marketers reporting billions of losses in 2020 alone. One of the key areas of criminal activity utilising 'paid-for' advertising business models involved targeted advertisements relating to deceptive products/financial schemes that involved well-known, trusted and reputable brands or individuals (Angeloni, 2022). Several studies demonstrated that UK consumers felt exposed and powerless in tackling scam ads and believed that online providers were not doing enough to support and protect them from harmful content online – a view that was shared by several expert witness testimonies during the online harms debate in Parliament (JC-1, JC-2, JC-3, JC-4, JC-9). Prior to the introduction of the Bill, the state appears to have taken on the burden of responsibility in this area, as the UK Advertising Standards Authority (ASA), the national advertising regulator, established its own Scam Ad Alert system. Introduced in 2020, the system provided a dedicated site/portal for UK consumers to report scam ads when spotted online, which would then lead to further cooperation with technology companies to take down scam ads (Advertising Standards Authority, 2020).

The building of scam ad mitigation into the Bill led to a fractious debate about the reallocation of responsibilities between the state, online service providers and individual citizens. The debate mirrored many others in the pre-legislative phase because it involved core issues of responsibility relating to actions required to mitigate harm. Ultimately, if the state or the private sector online service providers were unable or unwilling to block scam ads, then the consumer becomes the last line of defence for identifying and mitigating against harmful content – a requirement that would go against the core purpose of protecting individuals better against online harms that were difficult to detect by users. As such, allocating responsibility to end users would mean that individuals were deemed to have the capacity and capability to both identify and report scam ads, when there was a raft of evidence presented to the Committee that clearly indicated that was

Figure 3.2: Multiple actors with responsibilities



not the case. One of the key contentions arising from debate was that, despite the widespread harms scam ads cause, the business models deployed by the providers have paid-for advertisements at the core of service provision. Consequently, paid-for advertisements bring revenue to providers, whereas targeting harmful content, such as scam ads, is often regarded by providers as a regulatory cost. The Bill therefore sought to provide state influence by reallocating greater responsibilities to online service providers, while delegating a degree of latitude to preserve existing business models. The shifting responsibility for scam ad protection was reallocated from consumers to providers and it is this reallocation that gave rise to fractious debate and responsibility contention.

Findings

Our analysis of the pre-legislative debates revealed two core themes: (1) the responsibility contentions regarding the role of the state, online service providers and individuals regarding responsibility actions required to make the online space safer and (2) how digital responsibility manifests in private sector product design. Each theme is considered in greater detail next.

Responsibility contentions

The contentions that arise from the Bill debates analysis shows that there were different perspectives about which party bears responsibility for the implementation of technological systems to verify age of certain users and to prevent scam ads. The contentions are representative of the different responsibility actions required to implement solutions and the disagreements that flow regarding the allocation and discharge of responsibility. Responsibility contentions emerged in several thematic debates ranging from party willingness to absorb or avoid responsibility.

Throughout the debates, we identified situations where different parties were willing to accept or absorb responsibility for certain required actions and where the same or other parties sought to refuse or avoid responsibility for the technological implementation. As noted earlier, the debates highlight continuing attempts by the state and civil society groups for the Online Safety Bill to make providers more responsible for the publication of potentially harmful content on their platforms or for tackling online harms more broadly. The main legal vehicle for imposing greater responsibility through the Bill was the imposition of a general duty on providers to protect users from harmful content. Doing so would give rise to changing roles and relationships for state/regulator responsibilities and reducing the operational strategies of providers which placed greater responsibility for identifying and removing harmful content on individuals/consumers.

A number of expert witnesses and MPs/Lords argued that tackling harmful content was not a priority for providers, and that they were effectively abdicating responsibility for the

Figure 3.3: Putting profit first

safety of their users, even in situations where content breached their own rules and guidelines. For example, Imran Ahmed, founder and CEO of the Center for Countering Digital Hate, suggested that companies have their own ‘playbook’ in avoiding responsibility and were thus ‘incapable of regulating themselves’ and ‘put profit before people’ (JC-1). When asked why a lot of companies don’t do more work to target/remove harmful content, Ahmed suggested that if they did, ‘they would have moral culpability’ (JC-1). That said, even though it was identified that providers avoided responsibility, it was often unclear where and how responsibilities should be located. A panel of regulatory representatives from key regulators involved in harmful content moderation (Financial Conduct Authority (FCA), ASA and Competitions and Marketing Authority (CMA)) provided evidence to MPs that it was not clear where regulator and provider responsibilities should be allocated. It was not surprising, therefore, that responsibility allocation was a contentious issue.

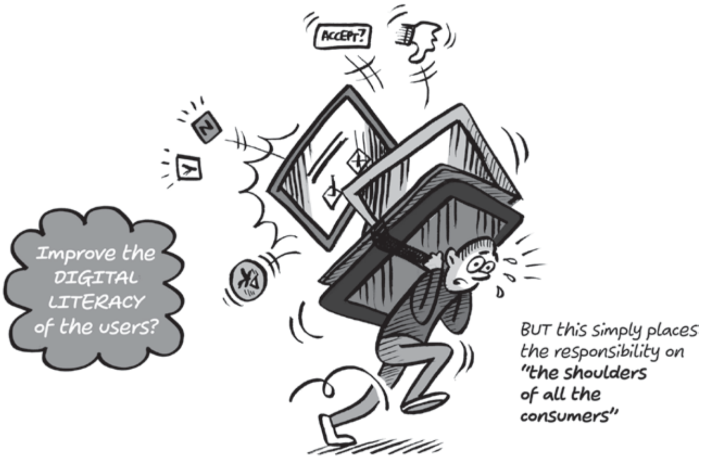
From the conflicts and tensions outlined so far, it is possible to identify a vacuum existing in the debates where several stakeholders were not entirely sure, or disagree, on where certain responsibilities should lie and with whom. These 'responsibility gaps' consisted of both legal considerations and societal/moralistic gaps, especially in relation to the actions of providers. In a legal/technical sense, contentions centred on the lack of regulation and corresponding security or safety controls that allowed providers to operate with relative impunity. On the latter point, one debate discussion noted how the Bill and online harms debate had highlighted a notable legislative gap for online content, especially in comparison to the way other media (such as TV or radio) are regulated (JC-9). For Clare Pelham, Chief Executive of the Epilepsy Society, the difference accentuated the isolation that many users experience online and how providers are 'completely beyond the reach of the law' (JC-4). Speaking to the claim that social media has become the 'Wild West online', these types of views demonstrate that there is little to zero accountability for online harm perpetrators operating on platforms nor the tech companies that allow perpetrators to function. In other words, everyone was in '[S]ome sort of badlands where there is no criminal law, no protection, no possibility of prosecution' (JC-4).

There was also clear contention and a responsibility gap about the protection of users' rights to freedom of expression and privacy in tackling harmful content online. This contention is of course not new and was a major consideration throughout the online harms debate. Concerns were raised about the Bill and the sweeping powers and responsibilities it gave to the state through Ofcom and the threat new regulatory powers posed to individual freedoms and rights, particularly in online spaces. The Joint Committee noted the Draft Bill 'provides duties on providers (and Ofcom) to have regard to the importance of the right to freedom and expression and privacy in various ways' (UK Parliament, 2022). While the Bill was supported by several organisations, particularly in the child protection community, several civil society organisations expressed major concerns about the Bill's scope and reach. The Internet Society claimed that 'the UK government was trying to legislate the impossible – a safe Internet without strong encryption – in the face of clear and consistent technical analysis saying that it cannot be done' (Internet Society, 2021). Similarly, Silkie Carlo, Director of Big Brother Watch, expressed concern during her Joint Committee testimony about the overarching regulatory powers that are tantamount to 'state-backed censorship of lawful communications' (JC-12). During her testimony, Carlo was challenged on this point by the Chair of the Joint Committee, Damian Collins MP. Collins took a different tack. Rather than solely focusing on the issue of whether content was unlawful or not, such as content that glamorises self-harm or anorexia under the UK's Equality Act, the more pertinent issue regarded 'to what extent companies are responsible for the environment they create and the harm that it could do' (JC-12). The role of the state and the protection of basic rights and freedoms was in itself unclear from a responsibility perspective.

Another key responsibility tension in the debates involved the notion of responsibilisation, whereby an individual (or group in society) is rendered responsible for a task or duty that previously would have been the role/responsibility of, say, the state or an associated institution.

In the context of the Bill and the wider online safety debate, responsibilisation is most contentious in relation to the expected role and responsibility of users. Responsibilisation contentions often surfaced in debates about incident reporting and the sense that citizens were being unduly relied on to report harms they encounter online. Individual users were thus being asked to bear the role of content moderation, reporting and policing that should be expected by both the state and providers (JC-9). To make individual users responsible for content moderation placed individuals in the situation of having the responsibility

Figure 3.4: The impact of responsibilisation on users


ACCEPT?
Improve the DIGITAL LITERACY of the users?
BUT this simply places the responsibility on "the shoulders of all the consumers"

of monitoring, reporting and protecting themselves from scam ads knowingly or unknowingly facilitated by provider business models and systems.

Related to responsibilisation, the issue of improving user digital/media literacy was also a persistent contention. Enhancing digital literacy of users will provide stronger self-protection mechanisms and will provide better choice mechanisms for users to identify and to not engage in disseminating harmful content online. Educative responsibility is partly borne by the state as it has been ascribed to Ofcom, though it still needs to be acknowledged that while the state carries some responsibility it still requires a significant requirement by individual users to (a) engage with such education, but also (b) put their improved literacy into practice.

Digital responsibility manifestations in product design

Running adjacent to contests about responsibilities and their allocation to ensure for online safety was an increased focus by the state on the business models of online service providers. To a certain extent, the greater focus on provider business models is not surprising given the whole policy purpose of law reform involved a recognition that online spaces needed to be regulated differently. Unsurprisingly, then, Joint Committee debates highlighted separations and differences between digital content as a potential source of harm to consumers and the requirement for the design, management and implementation of technology to ameliorate and protect consumers from such harms.

Like individual actions based on responsibilisation, the role of content moderation regarding information that could constitute 'harmful' was often used to shape discussions and surface tensions about responsibility allocation. Principally, debates about this issue concerned the targeted removal or takedown of harmful content, which has traditionally been the responsibility of online service providers who are often criticised for not taking enough action. A contentious part of the Bill debates regarded whether the regulator, Ofcom, should have more power to mandate companies to identify and remove content it deems to be harmful. Thus, the shifting role of the state emanated through the Bill's development. During Joint Committee witness testimonies, concerns were expressed about delegating too much implementation leeway to providers while other respondents believed that Parliament should set prescriptive requirements in legislation. The focus on potentially harmful content, and its management by providers, was therefore a specific source of consideration regarding product design. As noted earlier, the regulatory focus on product design as a core safety measure is one of the OSA's key features and one of its most controversial as it involves a change in responsibility by the state involving willingness to influence and shape market forces.

An ancillary debate to content management also emerged about how provider technologies are designed, managed and implemented in different contexts relevant to harmful content and business models. Moving beyond issues around the scope and effectiveness of content moderation, such debates focused on the various systems, processes and design features that have led to the creation and amplification of harmful content and activity online. Algorithmic accountability and platform recommendation systems were key considerations, and how such systems were designed to create profit through advertising and by keeping consumers online through strategies to attract and retain cognitive attention. Again, the state's willingness to consider this degree of involvement in shaping business model construction, and the actions that flow from core business operations, clearly signals a change in responsibility role and the actions required to mitigate broader, societal harms.

Contentious debates focused on whether and what responsibilities should be placed on service provider designers to ensure that they are made more responsible and accountable for the design and functionality of their platforms and the harms they could potentially create. Consequently, many of the most fractious elements of the debate related to the business models of providers and the potentially corrosive relationship between harmful content, consumers and profit motives. The debates highlight the fluctuations involving the different realms in which digital responsibilities are formed, namely the market, the state, the regulator and the individual.

Figure 3.5: Pre-set responsibilities in business models

THE
BUSINESS
MODEL
$
£
£
$
of tech companies has PRE-SET
the conditions for digital responsibility

Debates about the broader social consequences of provider business models became a key backdrop in which the role of the state and consumer with that of the structure of the market and the digital infrastructure that gives rise to online service provision were often played out. A key focus regarded the perceived imbalance for monetisation and profit at the expense of the safety and security of consumers. Provider business models were perceived as the root cause for many problems as they present themselves as 'free of charge' when they were really 'rooted in targeted advertising' based on maximising attention (JC-4). The crux of the issue is that, although online service providers have the technologies and capability to deal with the takedown/removal of most harmful content on their platforms (and already did so to a relative extent), it is argued by many commentators that they took a lax approach to tackling such content. Doing so may impact upon profits either through potential loss of advertising revenue or human resources costs to implement an effective moderation system.

In the Joint Committee hearings, this view was shared by MPs, academics, technologists and experts from charities/ campaign groups, who would often question what is stopping providers from doing more to tackle online harms and whether they were effectively 'profiteering from prejudice' (JC-2). Jimmy Wales (JC-6), the founder of Wikipedia, offered in his testimony a different business model that encompassed shared responsibilities across the platform. Under Wikipedia's business model, it was all users' responsibility to identify and remove misinformation. Content moderation was thus a shared responsibility of service provision and maintenance rather than a cost or threat to advertising-based business models. The juxtaposition of these two opposing business models reveals how content moderation responsibilities are realised and actioned to respond to online harms.

The criticism is well known to providers because it was previously used to highlight the lack of responsibility taken in relation to several key contentious activities involving potentially harmful content, including the proliferation of hate-speech, the live-streaming of terrorist content, algorithms promoting/linking to suicide content or the prevalence of scam ads that target vulnerable individuals. The Bill was consequently one of the first attempts to scrutinise and shape the dichotomy between provider made for profit business models and the safety of users. In doing so, the Bill would not only shape the business models of providers but also the responsibilities shaped the overarching digital market economy itself. Throughout the Bill development process, the question of whether providers would change their business models, and Parliament's ability to make them do so, was consistently ventilated. The Bill, in that sense, could set a precedent and send a message that the state was ready to govern the construction and operation of business model application.

Nevertheless, scepticism remained, with some expert witnesses suggesting that the Bill should go further and make providers, and company executives of providers, criminally liable for harmful content not removed from their platforms. The imposition of criminal measures would ascribe 'personal responsibility for senior managers in social media companies' and 'everything else will flow from it' (JC-4).

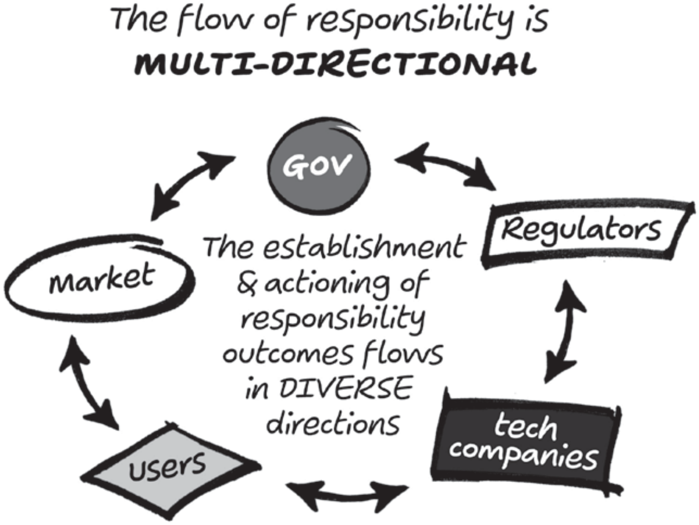
Thinking about digital responsibility

Our analysis indicates that the establishment and actioning of digital responsibility should not be assumed as axiomatic. During the debates, a responsibility consensus across key parties was not fully realised which meant that legal responsibility was established through the implementation of the OSA which commanded responsible action discharged through largely technological means rather than it being a consensus driven exercise. The Parliamentary debates thus demonstrate diverse types of responsibility action at play, which were mostly negative, particularly from the private sector parties, in trying to avoid responsibility for online harms and safety.

The debate contentions highlight that the to and fro of responsibility allocation is not unidirectional. Even though the OSA seems to indicate that once legal responsibility is established, through a Bill successfully completing its pre-legislative process and becoming a statute, it nonetheless still provides online service providers with avenues of interpretive leeway. Unidirectional application is thus not absolute, at least not in this situation. On its face, the pre-legislative debate history of the OSA presents an example of governmental activity taken where the private sector was unwilling to take responsibility for actions to create a safer environment by effectively mitigating potential content harms. The history, in one sense, could be viewed through this singular lens as purely an outcome of state action against unwilling regulatee response. Doing so, however, does not tell the full story. Instead, we contend that digital responsibility is multi-directional. The establishment and actioning of deemed responsible outcomes flowed in diverse directions. Some predictably flows from a command structure, such as the attempts by the state and civil society groups to make the private sector more responsible for their business models. Equally, on its face, the predictable actions taken by online service providers to avoid responsibility being allocated to them means that certain actions, such as reporting harmful content, become individual user responsibilities supported through technical processes and increasing educational capabilities. In this sense, the state also imposes some degree of responsibility on users by expecting them to educate themselves as a form of protection, especially from scam ads.

However, the process of establishing responsibility, and the actions produced, is dynamic rather than static. Take, for example, the civil society criticism regarding the OSA's potential for state-sponsored censorship. Consensus alignment takes place at various times, for example, when the state and civil society is aligned in critique of service provider business models. Yet, civil society and the state can be mis-aligned,

Figure 3.6: The flow of responsibility is multi-directional


The flow of responsibility is
MULTI-DIRECTIONAL
GOV
Regulators
tech companies
users
Market
The establishment & actioning of responsibility outcomes flows in DIVERSE directions

such as regarding the need for protection of fundamental rights or censorship concerns. Equally, the state and service providers find consensus alignment at certain points of the ongoing debate, such as the imposition of criminal penalties for provider senior managers, even though the proposal was largely supported by civil society. Again, these differences can be viewed as the normal cut and thrust of pre-legislative debate in which differences can be attributed to different notions of issue importance. Through a digital responsibility lens, the alignments and mis-alignments point to the different points of consensus that are required and generated when responsibility allocation is under discussion. Understanding the location points of consensus and recognising the actions that flow from established positions of alignment and mis-alignment are consequently important.

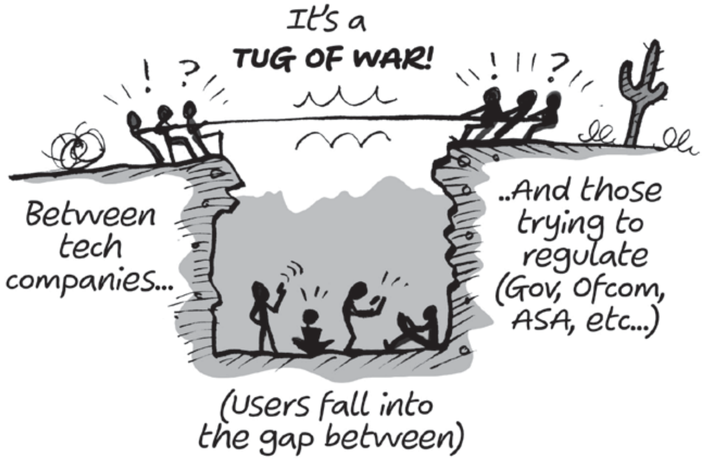
Our analysis of the debates showed the establishment of the OSA's differentiated duties of care, disrupted responsibility actions and expectations of which party should undertake desired regulatory actions. As noted earlier, even expert regulators with knowledge and experience of harms in the online sphere were concerned about where responsibilities could be located, including how state/provider responsibilities should be allocated. The situation is further complicated in an environment where the state attempts to impose responsibility, and the providers who would be imposed upon attempt to avoid responsibility by allocating actions and requirements to users. All of this shows that digital responsibility is transitory and flexible, which goes to explain why responsibility establishment and allocation of action are contentious issues.

The idea that responsibility establishment and actioning have a location consideration is important because it speaks to the importance of gaps, as highlighted earlier. The location of many debate contentions was the business model of service providers, the site where responsibility should be allocated and actioned but was not. Location is consequently important to consider because it highlights that responsibility is formed across locations and environments, which points to the multi-directional application of responsibility in the digital context.

In the OSA debates, the focus of business model contentions indicated that multiple parties thought regulation needed to be done differently and thus content moderation became the location focus of where regulatory activity needed to take place and the focus of responsibility allocation in terms of who should take reporting action, particularly the provider or user. The same could also be said regarding the regulatory gap and uncertain location involving the application of security controls. The imposition of controls was believed to be beyond the effective reach of law and responsibility establishment was therefore about the choice of appropriate technological measures which was best left to provider knowledge of needs. This also has much wider implications for the effectiveness of the security controls that form part of the digital design, for if the business model itself does not contain a framework that protects end users/consumers, then security controls will be less able to protect the end user/consumer. Unless the security goals for which controls are implemented are supported by an agreed distribution of responsibilities in the business model that promotes collaboration between the parties, the security controls will be limited in their effectiveness.

The idea of location and gaps also has a different dimension in the responsibility context. Responsibility gaps indicate consensus mis-alignments where responsibility allocation or actions fall into a lacuna and no party is responsible. The gaps themselves are multifaceted and can cross different environmental boundaries, such as public/private or legal/moral, which further diffuses the location and allocation of responsibility. Similarly, as highlighted earlier, a responsibilisation gap signifies situations where the state has retreated from undertaking certain responsibilities, thus creating a gap that needs to be filled. The latter highlights the shifting role of the state, including in the OSA debates. The state retreated from the management of online spaces and left

Figure 3.7: The sites of the state, the market and the individual


It's a TUG OF WAR!
Between tech companies...
..And those trying to regulate (Gov, Ofcom, ASA, etc...)
(Users fall into the gap between)

that activity to the market to establish responsibilities through individual terms of service. Yet, through the OSA, the state has returned to the construction of management expectations regarding safe online spaces and directly attempts to influence and shape market forces to create a safer environment by the prevention of harm. The OSA's focus on a safe online environment creates changing roles that the state is willing to establish responsibility and changing actions to ensure that establishment is fulfilled. Again, responsibility flows multi-directionally as state establishment and re-establishment flows across different realms and the state creates new ways to augment existing responsibility actions, such as takedown notices. Finally, the debates highlight a desire for shared responsibility. We read this desire as a signal for consensus where a gap has emerged. Accordingly, content moderation is not just to the action of taking down harmful content and it is more about the agreement of all parties to monitor and regulate across the digital environment.

Conclusion

Our analysis demonstrates how digital responsibility is established and realised at state-level. It highlights the digital policy challenge of tackling online harms as a prism to (a) examine the fractious policy debates to identify the responsibility contentions behind the debates and (b) identify how digital responsibility manifests itself within the different dimensions of digital product design, delivery and use by the private sector. Our overview of the online harms debate concluded that the business model of tech companies (and its relationship to the wider market economy) has pre-set the conditions for responsibility in this context. In other words, responsibilities in the context of the online harms debate are very much dictated and led by the business model of online service providers. For example, when considering how certain technologies are designed, managed and implemented on platforms, the conditionality shaping digital responsibility (be it a ‘duty of care’ or the protection of freedom of speech) is intrinsically linked to the market and business model of the platform. A few different tensions and conflicts arose between stakeholders which highlights that establishing responsibilities, even through state commanded forms of legislative action, such as the OSA, will not in itself obviate the need for consensus among core stakeholders.