Industrial and Corporate Change, 2021, **00**, 1–27
DOI: https://doi.org/10.1093/icc/dtab052
Original Article

Can self-regulation save digital platforms?

Michael A. Cusumano[1,*], Annabelle Gawer[2] and David B. Yoffie[3]

[1]Massachusetts Institute of Technology, 100 Main Street E60-300, Cambridge, MA 02142, USA. e-mail: cusumano@mit.edu, [2]Surrey Business School, University of Surrey, Rik Medlik Building, Guildford, Surrey GU2 7XH, UK. e-mail: a.gawer@surrey.ac.uk and [3]Strategy Unit, Harvard Business School, Boston, MA 02163, USA. e-mail: dyoffie@hbs.edu

*Main author for correspondence.

Abstract

This article explores some of the critical challenges facing self-regulation and the regulatory environment for digital platforms. We examine several historical examples of firms and industries that attempted self-regulation before the Internet. All dealt with similar challenges involving multiple market actors and potentially harmful content or bias in search results: movies and video games, radio and television advertising, and computerized airline reservation systems. We follow this historical discussion with examples of digital platforms in the Internet era that have proven problematic in similar ways, with growing calls for government intervention through sectoral regulation and content controls. We end with some general guidelines for when and how specific types of platform businesses might self-regulate more effectively. Although our sample is small and exploratory, the research suggests that a combination of self-regulation and credible threats of government regulation may yield the best results. We also note that effective self-regulation need not happen exclusively at the level of the firm. When it is in their collective self-interest, as occurred before the Internet era, coalitions of firms within the same market and with similar business models may agree to abide by a jointly accepted set of rules or codes of conduct.

JEL classification: L1, L2, L22, N00

1. Introduction

1.1 The argument

This article argues that recent events and current environmental conditions have created an opportunity for digital platforms such as social media and online commerce to engage in more self-regulation to avoid stringent government oversight. We also believe that the lack of clear rules regarding the operations of digital platforms has created, to some extent, a "moral hazard" (Arrow, 1963). Platform businesses have been able to use their power over users and other ecosystem members with relatively minor adverse consequences for themselves, incurring manageable financial penalties for violating competition and antitrust rules or digital privacy rules (Kira *et al.*, 2021). However, platform managers and boards of directors should have another motivation to curb their behavior: the potential for a broader "tragedy of the commons" (Hardin, 1968). This situation typically occurs when companies narrowly pursue their own self-interest, as in the case of a moral hazard, but then deplete an essential common resource that has enabled their prosperity. In the case of digital platforms, we see the essential common resource as user trust in a relatively open Internet market, which has become a global platform for digital commerce and information exchange. User trust, especially in dominant social platforms such as Facebook and Twitter, as well as in online marketplaces, has been declining over time (Khan, 2017; eMarketer, 2020). We begin with a literature review on self-regulation, followed by an exploration of the critical challenges facing digital platforms in 2021. We then examine several historical examples of firms and industries that attempted self-regulation before the Internet:

movies and video games, radio and television advertising, and computerized airline reservation systems. All dealt with similar challenges involving potentially harmful content or bias in search results. We follow this historical discussion with examples of digital platforms that have proven problematic in similar ways, with growing calls for government intervention through sectoral regulation and content controls. We end with guidelines for when and how specific types of platform businesses might self-regulate more effectively, in general and in the Internet era more specifically.

1.2 What are digital platforms?

Digital platforms bring together individuals and organizations so they can innovate or interact in ways not otherwise possible, using modern software, hardware, and networking technology (Cusumano *et al.*, 2019). Digital platforms aim to connect two or more market actors (market sides) and generate positive feedback loops among or across users in ways that bring increasing value to platform participants (network effects). In a global internet economy, where many resources are digital and utilized even if geographically dispersed, modern platform businesses are uniquely positioned to create and capture the value. They can facilitate both transactions (sales or exchange of products and services) and innovations (new features and products or services created by platform owners and third-party participants, usually called "complementors"). For example, marketplaces run by eBay, Amazon, Alibaba, and many other firms enable retailers as well as individual users to sell millions of items to buyers from around the world. Apple's iOS and Google's Android operating systems enable thousands of software developers to create millions of applications and services usable from mobile devices. As described in a broad and growing literature, digital platforms are, by design, central agents at the nexus of a network of value creators (see, for example: Gawer and Cusumano, 2002; Rochet and Tirole, 2003; Eisenmann *et al.*, 2006; Baldwin and Woodard, 2009; Gawer, 2014, 2020; Hagiu and Wright, 2015; Parker *et al.*, 2016).

A platform leader can capture a significant proportion of the value created in a distributed network, and it can monitor, control, and utilize ecosystem resources without owning them. The result is that platform businesses offer the potential for rapid, non-linear growth in profits, sales, and market value. Indeed, many of the most valuable companies in the world have built successful digital platforms along with market-leading products and services. Apple has captured roughly 90% of the world's smartphone profits, and Google ~80% of the smartphone operating system market and perhaps 90% of the global search market outside of China. Microsoft has captured 90% of PC operating systems and desktop productivity software. Google and Facebook control two-thirds of digital advertising. Amazon delivers roughly 40% of e-commerce goods in the United States and holds the largest share in cloud computing. Tencent and Alibaba have similarly dominant market shares in China's social media, messaging, and online marketplaces (Cusumano *et al.*, 2019).

Digital platforms can also be double-edged swords. They have generated trillions of dollars in wealth for their investors and provided many benefits to consumers, but they have also acquired outsized economic, social, and political power (Furman *et al.*, 2019; Gawer and Srnicek, 2021). Although any large firm may gain outsized market power, digital platforms, due to network effects and other factors, have been able to grow with astounding speed and sometimes win all or most of their markets (Eisenmann *et al.*, 2006). Their position of economic centrality, enhanced during the coronavirus disease 2019 (COVID-19) pandemic, has led to instances where some platforms have become strategic "bottlenecks" in the economy, as discussed in other articles in this special issue (Jacobides and Lianos, 2021; Jenny, 2021). Some scholars have argued that these bottlenecks and the potential for these markets to tip toward one big winner have occurred "unnaturally" (Marsden and Podszun, 2020). For example, unnatural tipping might happen when platforms in tight oligopolies hinder competitors via deliberate obstruction of multi-homing, making it difficult or expensive for end-users to use more than one platform at the same time or to switch from one platform to another. Moreover, during the past several years, we have seen instances of digital platforms disseminating fake news, distributing fake products, allowing content manipulation for political purposes, and bombarding

the public with misinformation on health and other essential matters. Some platforms have also mishandled business-to-business relations and engaged in pervasive, covert surveillance of digital users (Scott Morton *et al.*, 2019; Australian Competition and Consumer Commission (ACCC), 2019; Zuboff, 2019; U.S. Subcommittee on Antitrust, Commercial and Administrative Law of the Committee on the Judiciary, 2020).

These types of market behaviors require mediation by individual firms, industry coalitions, or government agencies (Boudreau and Hagiu, 2009). Some behaviors we can attribute to "market failures" such as abuses of monopolistic or oligopolistic positions or imperfect information. In reality, they are more akin to failures in platform and ecosystem management.

How and when platform companies choose to exercise or abuse their power, directly over their platforms or throughout their network of users, complementors, and other ecosystem participants, are thorny problems not covered well in the literature on platform strategy and management. The focus of most research has been on how to start and grow a platform business. In recent years, however, media commentators, legal scholars, and politicians have criticized social media platforms such as Facebook, Twitter, and Google YouTube for how they handle inappropriate content on their networks. eBay, Alibaba, and Amazon have constantly fought against fake reviews and the sale of counterfeit and sometimes illegal goods. Google Search, Apple, and Amazon have faced accusations of self-dealing, giving their own products and services preference over those of competing platform participants.

After decades of ignoring the potential harm from abuse or misuse of platform power, governments around the world have taken notice. The European Commission led the charge starting in 2010 with its first antitrust lawsuits against Google, followed by two other actions in 2016 and 2017. Attempts to reign in big tech platforms in the United States emerged later than Europe: in the fall of 2020, the US government launched aggressive antitrust actions against Google and Facebook. Expanded regulatory actions were arising in the European Union, Germany, Australia, Japan, Russia, and China in 2021 (Mozur *et al.*, 2021). It is reasonable to expect that the insurrection at the US Capitol on 6 January 2021 could accelerate government actions or at least increase scrutiny of social media platforms globally.

Ironically, while the reputation of some digital platforms has clearly suffered, their revenues and profits have continued to soar, especially during the COVID-19 pandemic, when so many activities shifted to the Internet. As a result, despite a growing lack of trustworthiness, the implications for their businesses in the future remain uncertain. We may mistrust digital platforms, but we must still rely on them. In this context, there has been growing pressure for government intervention, which is likely to vary by country. Due to their complex technology and rapidly evolving features and services, it may be difficult for governments to intervene effectively in the operations of digital platforms. Poorly designed interventions also could generate harmful side effects for consumers, ecosystem participants, and the global digital economy (Petit and Teece, 2021).

For example, the banking industry faced a similar crisis of trust and ineffective policy decisions during 2008–2009. In the 1990s and 2000s, the Clinton and Bush administrations eliminated most regulations for banks, subprime mortgages, and related investment instruments. Banks then aggressively took advantage of deregulation and nearly destroyed trust in their common resource—the global financial system—with virtually no efforts to self-regulate. This case exhibited features of both moral hazard and a tragedy of the commons, resulting in a damaged financial system that took years to repair. We should have a similar concern for digital platforms and the platform economy in the 2020s. Trust is fundamental to a well-functioning World Wide Web. If companies and users that rely on the Internet destroy trust in that platform, then the utility of the Internet could be greatly diminished. Going forward, we need to understand *when* it is in the interest of digital platforms to self-regulate aggressively in order to preserve the system in which they operate, and *how* these firms can effectively curb their own behavior or limit abuse by users and other ecosystem participants.

We believe that more effective self-regulation, facilitated by governmental pressure, may hold the key to avoiding a potential tragedy of the commons. While some executives, such as Facebook's Mark Zuckerberg, want governments to resolve these challenges for them, the complexity and speed of technological innovation make it doubtful that government regulation, alone, will

suffice. Our sample is small and exploratory, but the research suggests that a *combination* of self-regulation and credible threats of government regulation may yield better outcomes than government invention alone. In addition, effective self-regulation need not happen exclusively at the level of the firm, and there may well be a "gray zone" in between formal government regulation via legislation or other means and informal self-regulation by companies. For example, history suggests that, when it is in their collective self-interest, coalitions of firms in the same markets and with similar business models may agree to abide by a jointly accepted set of rules or codes of conduct. Indeed, we identify several cases before and after the Internet era where coalitions of firms, working in their collective self- interest, have worked together effectively to regulate use of their platforms, products, and services.

2. Self-regulation and the regulatory environment

2.1 Definitions and motivations

What do we mean by self-regulation? This term generally refers to the steps companies or industry organizations take to preempt or supplement governmental rules and guidelines that govern their activities. We usually contrast self-regulation with government regulation that can take a direct form such as legislation and penalties for violations or an indirect form such as taxes, subsidies, permits, licenses, and similar measures. At the individual firm level, self-regulation ranges from a firm monitoring its own possible regulatory violations to proactive corporate social responsibility initiatives that aim to contribute to the community as well as improve the firm's public image. Industry-level self-regulation typically involves forming collective institutions like industry associations that perform regulatory functions, such as granting licenses and permits, often in collaboration with government agencies (Maitland, 1985: 135).

Although we can define self-regulation in general terms, scholars are not always in agreement on the specific details (Sinclair, 1997; Gunningham and Rees, 1997; Freeman, 2000). Some describe self-regulation as any rule imposed by a non-governmental actor, while other scholars point to any rule created and enforced by the regulated entity itself. In this paper, we follow Coglianese and Mendelson (2010). They define self-regulation as "any system of regulation in which the regulatory target—either at the individual-firm level or sometimes through an industry association that represents targets—imposes commands and consequences upon itself" (Coglianese and Mendelson, 2010). They also define "meta-regulation" as related but distinct from self-regulation in that it refers to ways outside regulators deliberately, rather than unintentionally, seek to induce targets to develop their own internal, self-regulatory responses to public problems. This notion of meta-regulation again suggests a blurring of the distinctions between explicit and informal government influence on firm behavior.

Our use of the term self-regulation encompasses both self-regulation and meta-regulation. This approach offers the advantage of not being restrictive regarding either the nature of the rules or their consequences. Our definition simply says that self-regulation occurs when the "regulator" issues commands that apply to itself. The "regulator" here does not have to be a government agency; it can be a company or an association of firms. We consider an action to be "self-regulatory" even if motivated by implicit threats from outside regulators, whether or not the outside threats express intentional efforts to encourage self-regulation. We can see this type of self-regulation in a variety of contexts. For example, there is the Motion Picture Association (MPA) of America's movie rating system (Campbell, 1999), the Forest Stewardship Council's Sustainable Forest Initiative (Gunningham and Rees, 1997; Nash and Ehrenfeld, 1997; von Mirbach, 1999), and the guidelines of the Entertainment Software Rating Board (Laczniak *et al.*, 2017). Other examples include the International Organization for Standardization's guidelines for environmental management systems and the American Petroleum Industry's Strategies for Today's Environmental Partnerships, also known as STEP (Nash and Ehrenfeld, 1997).

Some scholars believe that self-regulation can be faster, better, and cheaper for consumers as well as for companies (Gunningham and Rees, 1997: 366). In effect, self-regulation usually shifts discretion over how to regulate from the regulator to the target. This can be beneficial because targets are likely to have greater knowledge of and more information about their own operations. The targets may also perceive rules of their own making to be more reasonable

than those imposed by outside entities, and this may increase the likelihood of compliance (Coglianese and Mendelson, 2010). Self-regulation and meta-regulation may also allow regulators to address problems when they lack the resources or information needed to craft sound discretion-limiting rules, which can happen when a regulatory problem is complex or an industry is heterogeneous or changing quickly. It is true that companies might well prefer no regulation, especially if curbing certain practices may be harmful to their revenues or profits. Fundamentally, self-regulation should come into play when it is in the best interest of firms to modify user behavior in order to promote a healthier industry long-term. That is, self-regulation should work when it "promises simultaneously to allay business fears of further government encroachment and to restore the public's faith in business. What is more, it asks of a business only that it behaves in its own enlightened self-interest" (Maitland, 1985).

At the same time, we can find in the literature acknowledgements that self-regulation is no panacea for all problems that might occur in a market. The primary limitation of self-regulation involves incentives. Although companies "may have better information to find solutions to public problems, they may lack the incentives to do so" (Coglianese and Mendelson, 2010). However, there is still the issue of possible divergence between the firm's private interests and society's interests. Leaving companies to police their own behavior can devolve into what some scholars have termed a regulatory "charade" (Gunningham and Rees, 1997: 366).

Critics of self-regulation also suggest that proponents underestimate the limited discretion managers actually have in a competitive economy to undertake costly actions for the public good because this would render them vulnerable to rivals. For example, Maitland (1985) argued that an individual firm's interests as a competitor in the marketplace often diverge from its interests as a part of the wider society (or, for that matter, as a part of the business community). Regardless of how well-intentioned they are, firms engaging in self-regulation can be seen as investing in a collective good. Therefore, they are subject to the typical problems plaguing collective action. As individual firms typically cannot control how competitors will react, they would rationally fear that competitors would use this for their advantage, leading to the target firm refraining from engaging in self-regulation. Sometimes, the problem is that it is rational for each firm to put its own individual interests ahead of collective interests. At other times, the problem is to coordinate firms' expectations regarding their fair shares of the market.

The literature offers some tentative insights into conditions under which self-regulation can work particularly well. The size of the industry and whether it consists of homogenous actors appear to matter. Shared perceptions of a common interest, and the ability to monitor actions and implement sanctions, play an important role as well. For example, both the chemical industry and the nuclear industry self-regulation initiatives were spurred by accidents that brought intense public scrutiny. The Responsible Care initiative, launched in the mid-1980s by the Chemical Manufacturers Association in response to the Union Carbide chemical accident in Bhopal, India, led chemical companies to make commitments to upload certain environmental, health, and safety values. Empirical studies deemed that it mostly failed to deliver substantial environmental improvements (Rees, 1994; Nash and Ehrenfeld, 1997; King and Lenox, 2000). By contrast, the Institute of Nuclear Power Operations (INPO), established as an industry association in response to the Three Mile Island nuclear accident in 1979, appeared to be successful in improving nuclear plant safety (Russ, 1982). Analysts attributed this success to the nuclear power industry being much smaller and more homogeneous than the chemical industry. Individual firms shared more common interests and a common business model, which made it somewhat easier for leadership at the power companies to rein in potential outlier firms (King and Lenox, 2000). By contrast, the chemical industry contained a few large firms deeply concerned about their consumer image and, correspondingly, the image of the industry as a whole. But the industry also contained many smaller chemical firms with much less of a stake in the sector's overall reputation (Gunningham, 1995).

Self-regulation's primary weakness is the potential absence of a mechanism to align targets' incentives and interests. As a result, industry-specific or trade-association codes of conduct often face collective action problems in their implementation and enforcement. The efficacy of self-regulation can sometimes depend on the degree to which some outside threat reinforces voluntary, collective efforts at self-control (Segerson and Miceli, 1999). Firms in a large, heterogeneous

industry can probably defect more easily on any self-regulatory collective action (Olson, 1965). In the case of the chemical industry, the defection problem may have been exacerbated by Responsible Care's weak monitoring and transparency mechanisms, at least during its first decade. INPO may have proven more successful simply because it regulated fewer firms and firms with tighter interdependencies. It also benefited from the Nuclear Regulatory Commission (NRC), which served as a kind of "regulatory 'gorilla in the closet' that could provide sanctions for opportunism and act as an outside auditor for the program" (Rees, 1997: 477). The NRC stood as a more credible, even if implicit, threat to the nuclear power industry than did the divided and contested governmental authority of Occupational Safety and Health Administration and Environmental Protection Agency (OSHA and EPA) over the chemical industry (Coglianese and Mendelson, 2010).

In short, self-regulation appears to succeed when targets decide it is in their best interest not to defect from self-imposed (or sectoral-imposed) standards. If compliance is costly, firms may need some kind of external pressure to provide an incentive for voluntary adherence. External pressures can be economic, social, or regulatory (Gunningham *et al.*, 2003). Self-regulation by an industry group also seems to work best when the sector consists of a small, relatively homogeneous number of interconnected or interdependent actors and face shared challenges. In addition, effective self-regulation seems to require firms capable of monitoring themselves as well as exercising credible controls on their own behavior. External factors such as the threat of direct government regulation can add additional important incentives for firms to self-regulate.

2.2 The changing regulatory environment for digital platforms

Part of the motivation to regulate digital platforms reflects classical economic concerns. In the non-digital world, governments traditionally addressed the abuse of economic power through antitrust laws, such as in the famous breakups of Standard Oil and AT&T. We also see monopolistic behavior in the digital economy and powerful network effects driving some winner-take-all or winner-take-most outcomes (Parker *et al.*, 2021). Indeed, every large platform company has faced American, European, or Chinese antitrust actions. However, antitrust actions are lengthy and costly for all concerned. Moreover, when it comes to digital platforms, antitrust is not the only issue prompting calls for more government intervention or self-regulatory actions, as we observed in the role Facebook and Cambridge Analytica played in the election of President Donald Trump in 2016 (Cusumano *et al.*, 2019: 6–8).

Digital platforms have also benefited from exceptionally favorable or vague regulations. For example, social media platforms have done particularly well under Section 230 of the United States 1996 Communications Decency Act. This legislation states that, "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider" (U.S. Congress, 1996). Initially created before the advent of modern social media platforms, the law gave online intermediaries broad immunity from liability for user-generated content posted on their sites. In 2020, however, the CEOs of Facebook, Google, and Twitter, as well as a coalition of smaller platforms led by Snap, Reddit, and Tripadvisor, all agreed on the need to modify Section 230. They foresaw that change is coming and prefer to shape those changes rather than have new regulations imposed on them. Proposed changes from the companies would clarify how they moderate online content and allow users to appeal decisions that platforms make or give users more control over algorithms sorting their content. Some recommended changes would make it more difficult for platforms to use Section 230 as protection in lawsuits if they are "intentionally facilitating illegal activity" (McCabe, 2020).

Revising Section 230 is not the only option. Free market advocates have contended that competition will eventually eliminate problematic practices, as new firms enter with different approaches (Petit and Teece, 2021; Teece and Kahwaty, 2021). Others disagree, claiming that winner-take-all-or-most outcomes make it extremely difficult to dislodge dominant platforms. The latter advocate for more robust regulation to make sure markets impacted by digital platforms remain competitive (Khan, 2017; Australian Competition and Consumer Commission (ACCC), 2019; U.S. Subcommittee on Antitrust, Commercial and Administrative Law of the Committee on the Judiciary, 2020).

Although digital platforms have the potential to abuse their considerable power, they also can limit their own influence through governance rules that shape behavior in their broader ecosystems of users and partners. Platform leaders can adopt more versus less "open" technical standards by themselves or consortium partners, such as for operating systems and application programming interfaces (Jansen and Cusumano, 2013). They can institute governance rules determining who can and cannot access the platform and access points such as app stores, and influence what users or complementors can do on the platform (Gawer and Cusumano, 2002; Parker *et al.*, 2016; Cusumano *et al.*, 2019). Indeed, governance rules may even put digital platforms in the position as "private regulators" (Boudreau and Hagiu, 2009), with the ability to shape broad industry architecture as well as the behavior of platform participants (Jacobides *et al.*, 2006; Jacobides and Lianno, 2021).

Digital platforms and their ecosystems constitute both a new organizational form and a new challenge for government regulation. However, arguments in favor of reigning in the behavior of powerful corporations are not new. History teaches us that debates over regulation are likely to emerge whenever new industries and dominant firms emerge, and the line between regulation and self-regulation is often blurry. Consequently, reviewing some relevant historical cases should help us understand how self-regulation has evolved and worked or not worked in the past, without the complications added by the Internet and digital technology. We may also get a better view of how self-regulation and government intervention can function together. Since scholars have studied antitrust more extensively than regulation of social media, messaging platforms, and digital search engines, we selected historical cases that related primarily to content, advertising, and search self-preferencing. We will review these same issues when we return to our discussion of digital platforms in the Internet era.

3. Pre-internet: self-regulation before the advent of digital platforms

This section examines firm and industry efforts to self-regulate in several industries that predated the Internet age. We explore content in movies, video games, radio and television broadcasting, advertising over radio and television, and computerized airline reservation systems used by multiple firms. This is not a random sample of companies and industries. Rather, we choose a few examples that might prove to be instructive. Some of these industries already had some regulation, such as licenses for radio and television broadcasting frequencies or licenses to run commercial airlines. However, each company and industry faced the threat of additional government intervention into specific non-regulated activities, which in turn triggered self-regulation.

Although there are many differences between these historical examples and modern digital platforms, the similarities are important and instructive. Movies, video games, and broadcasting all faced issues that revolved around the appropriateness of content in ways that resemble today's debates about social media and messaging platforms. For example, the movie and videogame industries resorted to self-imposed and self-monitored rating systems, which still operate in 2021, to keep regulators at bay. The radio and TV broadcasting sector faced pushback on the appropriateness of advertisements, resembling what we see today in online advertising and digital content. Finally, airlines introduced self-preferencing in search results for reservations, which is similar to complaints made against Google Search and Amazon Marketplace's search engine. Firms in each of these sectors made multiple attempts at self- regulation; some succeeded, others failed.

The historical examples we selected also fit descriptions in the literature of when self-regulation is more likely to succeed than not and resemble market dynamics that we see in the Internet age. Firms in each industry were relatively homogenous, with common interests and business models that negative public perceptions could greatly impact. There were some large competitors but no single dominant firm that might have ignored attempts at industry-level self-regulation or government intervention. There were relatively few competitors in each market, and similar business models, both of which increased the chances of reaching a consensus on self-regulation in order to prevent or curb government regulation. These firms were in relatively new industries during their heyday and most likely understood the new technologies (cinema, radio and television broadcasting, video games, and computer reservation systems) better than government legislatures or regulatory agencies.

There were also platform-like elements (i.e., multi-sided markets and network effects) in these historical cases, even without the Internet. For example, movie theatres showed content that came from different production companies and studios. Radio and television broadcasters showed programming from their own studios as well as from outside providers. The more theatres or radio and television sets diffused through the economy, the more valuable these content delivery channels (or platforms) became for showing both content and advertisements. Video games also depended heavily on content from multiple sources and grew in value as more users bought consoles. Airline reservation systems evolved to handle content from multiple companies in order to facilitate a well-functioning travel industry, with more and more flights increasingly accessible to the public.

Finally, history teaches us that failure to regulate or self-regulate can have serious or even catastrophic consequences for an industry or the public good. For example, the data on cigarette consumption and cancer rates were highly correlated with the lack of regulation or self-regulation in cigarette advertising. As soon as advertising was banned, cancer rates and consumption declined (Jha, 2020).

3.1 Movies and video games: policing and classifying content

For centuries, governments and other institutions have used censorship to protect people, particularly children, from being corrupted by dangerous ideas and images. Regulation and self-regulation of film in the United States is a case in point. In 1907, Chicago enacted the US' first film censorship law, which granted the police commissioner power to ban theaters and penny arcades in the city from exhibiting "obscene or immoral" material. Movie operators challenged the ordinance but lost (Lupton, 2018). In a unanimous 1915 decision, the US Supreme Court upheld a similar censorship statute enacted in Ohio (Mutual Film Corporation v. Industrial Commission of Ohio, 1915). The decision stood as precedent for the next several decades (Joseph Burstyn, Inc. v.Wilson, 1952).

Ordinances similar to those in Chicago and Ohio appeared in other American cities and states, creating a complex and unpredictable regulatory environment (Wittern-Keller, 2008: 36). In 1922, industry participants founded the Motion Picture Producers and Distributors of America (MPPDA), later named the MPA, largely to forestall further censorship and to supplant the patchwork system (Piepenburg, 2018: 101–102). The National Board of Review, an independent organization financed by industry members, had been reviewing and approving movies since 1909, but its standards were less strict than those of many censor boards and pro-censorship advocates (Wittern-Keller, 2008: 23–24). Under the leadership of former postmaster general Will H. Hays, the MPA attempted to be more effective, basing its first list of "don'ts" and "be carefuls" on what was "habitually condemned by censoring boards" (Piepenburg, 2018: 101).

Movie producers largely dismissed industry guidelines as bad for business. Meanwhile, Congress considered federal censorship legislation and over one hundred bills for stricter regulations appeared in 44 U.S. States (Couvares, 1992: 592–593). With this looming regulatory threat, in combination with pressure from Catholic organizations, the MPA finally was able to exert more influence over the industry. In collaboration with Father Daniel A. Lord, a Catholic priest, the MPA developed a more stringent set of guidelines in 1930 called the Motion Picture Production Code (also known as the Hays Code). There was loose enforcement until the Catholic Legion of Decency made threats to call for film boycotts. Soon after, the Production Code Administration (PCA), the MPA's censorship branch, decided to levy $25,000 fines against studios for code violations (Conant, 1978: 40). By 1945, 95% of all US films were made with oversight from the PCA, and movies from that period were, on the whole, significantly "cleaner" than those released before (Piepenburg, 2018: 104). Life became easier for distributors and exhibitors: between 1934 and 1940, "problem films" requiring cuts or bans declined by 70%, according to New York's censorship authority (Wittern-Keller, 2008: 294). Self-restraint became part of how companies produced movies. When the Supreme Court declared New York's censorship statute null and void in 1965, there was little uproar or press coverage (Wittern-Keller, 2008: 253). Three years later, the MPA officially abandoned the Hays code (Hunt, 2018).

To the extent that the public cared about regulating the movie industry, the concern shifted to age classification and restrictions. The same year it abandoned the Hays code, the MPA formed the Code and Rating Administration (CARA), which assigned five ratings: G for "general audiences;" PG for "parental guidance suggested;" PG-13 for "parents strongly cautioned;" R for "restricted, no one under the age of 16 admitted without a parent or guardian;" and NC-17 for "no one 17 and under admitted." Since 1968, CARA has based its decisions on the recommendations of a panel of paid parent raters with children between 5 and 15 years of age. Although it has made the rough guidelines for its ratings public, the full rating process remains secret (Piepenburg, 2018: 111).

The MPA movie ratings figure heavily into financial performance. G-rated movies are generally less successful than movies with more restrictive ratings because the G rating carries the connotation of being 'too kiddie.' On the other end of the spectrum, NC-17 movies tend to be less successful because their audience is restricted, and studios often make changes post-production to pull a movie back into R-rated territory (Bailey, 2011).

The "happy medium" for movies is PG-13: high enough that it can feature sufficient violence and (to a lesser extent) sex to draw audiences, but low enough that no audience restrictions apply. PG-13 movies are by far the industry's largest source of revenue, accounting for 48% of box office earnings between 1995 and 2020 (Nash Information Services, 2021). However, this self-regulatory system still has flaws. Over the past few decades, MPA has received repeated criticism for letting studios drive up sales by marketing violence to children (Piepenburg, 2018: 114). PG-13 movies have gradually become more and more violent: in fact, one study found that PG-13 movies now feature more gun violence than R movies (Sorrel, 2017). Another study found that, relative to government regulators in other countries, the MPA "assigns more lenient ratings to violent movies with high teenage appeal," a pattern consistent with "a strong preference for maximizing industry revenue (Lampe and McRae, 2014)."

Parent groups also became concerned with regulating video games, much in the same way as we saw with movies. In the video-games case, self-regulation emerged due to external pressure from politicians and parents. In particular, as new, more violent video games became popular, parents worried they would make children antisocial and violent, and they created substantial political pressure to regulate the industry. At a hearing in 1993, a senator warned: "If the video game industry cannot effectively police itself, then Congress will have to" (Puga, 1993). Congressman Joseph Lieberman introduced legislation for a government-mandated commission to establish ratings and warned the industry that he would push it through unless the industry voluntarily did it themselves. Facing a serious threat of government intervention, the video game companies created the Entertainment Software Ratings Board (ESRB) that promulgated a voluntary but a standardized rating system for video games (Kohler, 2009). Today, almost all video games sold in the United States and Canada are voluntarily rated by the ESRB (Piepenburg, 2018: 118).

Originally, to receive a rating, video-game makers had to pay $500 and submit a 30 minute selection of the most graphic scenes in their games to the ESRB, where a small team of adults from the New York area reviewed the videotape (Corcoran, 1994). The basic process remains roughly the same today and resembles the rating system for movies: EC for "early childhood," E for "everyone," T for "teen," M for "mature" and AO for "adults only." In addition to rating games, the ESRB issues and enforces guidelines on marketing and selling video games. Because M and AO games are age-restricted, ESRB conducts regular audits of retailers to ensure that they verify customer ages before selling M and AO-rated games (Entertainment Software Rating Board, 2021). The ESRB may even impose sanctions, including fines of up to $1 million, on companies that submit incomplete or misleading materials for review (Piepenburg, 2018: 119).

Politicians and advocacy groups have repeatedly called for more government regulation of video games, but regulatory experts and agencies both have praised the ESRB for its effectiveness (PR Newswire, 1999). Early on, one senator said that he was "particularly pleased" by the ESRB's system and thought it was "the most comprehensive rating system in existence for any entertainment"'(Corcoran, 1994). In 2009, the Federal Trade Commission (FTC) concluded in a report to Congress that the video game industry had "the strongest self-regulatory code" of all

three entertainment industries it had investigated, the other two being music and video (Federal Trade Commission, 2009).

Movies and video games illustrate that self-regulation can indeed work well for certain types of problems and under certain conditions. The type of problem that this self-regulation addressed was the shared perception among the public and politicians that content needed policing. Government action threatened the entire industry, and firms managed to address the threat as a group, without seriously undermining their business models. Self-regulation would not likely have emerged without mounting pressure from government actors and public groups. In particular, the prospect of more intrusive government regulation led to an industry-level agreement on a rating system and credible self- enforcement, with the ability to impose fines.

3.2 Broadcasting and advertisements: restricting harmful content

In 1934, the US government created the Federal Communications Commission (FCC) to introduce a licensing system and allot channels in the electromagnetic spectrum to television and radio broadcasters. The FCC also used its authority to regulate the content of radio and television shows by banning licenses to stations' broadcasting content it believed to be contrary to the "public interest, convenience, or necessity" (U.S. Congress, 1934). Major examples of influential FCC policies include the Fairness Doctrine, which required broadcasters to provide balanced coverage of controversial issues (eliminated in 2011), and the Equal Time Rule, which requires that broadcasters provide political candidates with the same opportunities to advertise and campaign.

The National Association of Broadcasters (NAB) adopted its first "Code of Ethics" for radio in 1928. Without the ability to enforce it, however, the Code "lapsed into obscurity" (Jaramillo, 2018: 27). In 1946, the FCC released the "Blue Book," which set guidelines for how broadcasters could meet the public interest. The NAB responded by arguing that the Blue Book violated freedom of speech. Following an investigation by a House select committee that recommended the FCC abandon it, the Blue Book also went on to have virtually no policy impact. It did, however, make the threat of increased regulation more apparent and marked a turning point in the industry's path toward stricter self-regulation (Jaramillo, 2018: 29).

Another milestone was the Benton Bill, named for Senator William Benton. In the early 1950s, Benton advocated for the creation of an "annual Blue Book" to review the content of educational and public service programming (Jaramillo, 2018: 150). In response, the newly named National Association of Radio and Television Broadcasters (NARTB) began work on a set of standards for television programming (Jaramillo, 2018: 162). The NARTB Television Code came into effect in 1952. The idea was to respond to the major concerns of the public at the time, which were the content of children's programs, advertising for alcoholic beverages, obscenity, and profanity (Sezey, 1952: 15). Unlike the NAB's first Code of Ethics, the Television Code set specific standards concerning everything from the depiction of gambling to advertisements for fortune-telling services.

The Television Code also included a "fake news ban": It warned broadcasters against "present[ing] fictional events or other non-news material as authentic news" and broadcasting "dramatizations in any program which would give the false impression that [it] constitutes news" (National Association of Radio and Television Broadcasters, 1952: 4). Over the next four years, the NARTB responded to 1663 letters of viewer disapproval. Only a small percentage of these letters labeled content "immoral" or "indecent." Instead, the most common complaint—94% of all letters—concerned television advertising content and frequency (Pondillo, 2010: 187). Despite these complaints, the advertising industry never paid much heed to the NARTB Television Code or the internal codes of broadcasters like the NBC (Pondillo, 2010: 188). The responsibility to regulate fell to the broadcasters. In keeping with the Television Code, they barred themselves from running ads that were either in "poor taste" or that included misleading phrases like "free" and "limited time offer" (National Association of Radio and Television Broadcasters, 1954: 32). One notable exception was alcohol advertising. The liquor industry voluntarily stopped running ads for whiskey and other hard liquors in 1936. Ads for beer and wine followed their own code of ethics, which barred the depiction of the act of drinking, "amorous activity as a result of consuming beer," or Santa Claus in any alcohol-related commercial (Pondillo, 2010: 185).

Cigarettes also became popular in large part because of advertising, beginning with Camel's first print ad campaign in 1913. By the 1930s, cigarettes were by far the most popular tobacco product in the United States. Advertising fueled a sharp rise in per-capita consumption of tobacco between 1940 and the mid-1950s (Burns *et al.*, 1997). At the peak of their popularity in 1953, 47% of American adults smoked cigarettes (Cummings and Proctor, 2014). Medical studies suggesting a link between smoking and lung cancer began to appear in the 1920s (Gardner and Brandt, 2006: 222). Yet not until 1957, after several decades of increasing rates of lung cancer and growing concern among doctors, did the United States Public Health Service release its first official statement linking smoking to lung cancer (Cummings and Proctor, 2014).

In 1962, NARTB President LeRoy Collins called for the control of tobacco advertising on the grounds that it was hazardous to the health of children. He claimed that the moral responsibility lay with tobacco manufacturers, advertising agencies, the "sports figures who permit their hero status to be prostituted," and broadcasters. The industry did not welcome Collins' call for moral responsibility (Gould, 1962: 138). The industry resisted cutting cigarette ads until the Surgeon General released a report in 1964 stating definitively that cigarettes caused lung cancer and chronic bronchitis. Still, few advertising agencies felt any qualms about working with cigarette companies (Pollay, 1994: 138).

Cigarette companies did respond to Collins' criticism; however, the six major firms dropped all advertising from radio, magazines, and newspapers, leaving only television. But they also created the "Tobacco Institute" to serve as a trade organization and promulgate suggestions about advertising. Among its suggestions was to avoid sponsoring programs aimed at young audiences. The major firms all adopted the Institute's suggestions although they had a minimal impact on advertising practices: children and young adults were 25% of the audience for cigarette ads and the average teenager saw over 100 cigarette ads per month. The next attempt to self-regulate cigarette ads came via the NARTB, which asked that shows for younger audiences not include advertisements depicting smoking, but their efforts were ineffective (Pollay, 1994: 138–139).

In 1967, the FCC granted an application of the Fairness Doctrine to cigarette advertising and public announcements about the health risks of cigarettes began to appear on air. The tobacco industry took the FCC all the way to the Supreme Court and ultimately lost after the Court refused to review the case. Around the same time, the FTC announced plans to require warnings on cigarette ads, and the FCC announced it was considering banning cigarette ads altogether. Seeing that the end was near, the chairman of the Tobacco Institute agreed that all tobacco companies would cease broadcast advertising if the institute were granted immunity from antitrust regulation. Congress went ahead and passed a law banning cigarette ads anyway and advertising for cigarettes disappeared from broadcast television in 1971 (Pollay, 1994: 141).

The lessons for regulation and self-regulation in broadcasting and advertising provide insights that complement what we have learned from movies and video games. The main difference between these two sectors is that, in broadcasting and advertising, any regulation or self-regulation on cigarette ads would negatively affect sales of cigarettes, whereas ratings of movies and video games seemed to have less of a negative impact on sales. There is a striking conflict with the economic incentives of all players, which might be difficult to reconcile unless industry leaders believed that more serious government actions (such as an all-out ban of cigarettes) were possible. Government intervention was much more aggressive than in other sectors, probably because of the obvious threat to public health. However, when faced with even more intrusive government action tied to protecting public health, tobacco firms were willing to make dramatic changes in their behavior.

3.3 Computerized airline reservations: search and self-preferencing

Airlines were early adopters of computers for internal operations and eventually for passenger reservations and ticketing. In 1954, years before launching the now-famous Sabre computerized reservation system (CRS), American Airlines began working with IBM to use computers to manage revenue accounting and seat inventory. At the time, American handled passenger reservations through an intensive manual process called "book to board." This required

employees to send bookings via teletype back and forth between reservations offices. The biggest technological hurdle was finding a way to process passenger records, an enormous amount of data, in real-time. Overdue and over-budget, Sabre became operational in 1964, and this system and competing CRSs became commonplace in travel agencies by the 1970s (Head, 2002).

In the United States, the federal government had been regulating commercial aircraft flights since the 1920s (Federal Aviation Authority, 2017). However, airline reservation systems remain unregulated, even today. Moreover, the early CRSs only showed flights from the airlines that ran them. American, United, and a few smaller competitors began to permit other airlines to list flights on their CRSs in the late 1970s (Locke, 1989: 219). However, all reservation systems exhibited bias: a CRS usually showed flights first from the airline that owned the system, and it was often difficult to find flights from competing airlines (Salpukas, 1982). This practice came to be known as "screen bias." Today we would call this "self-preferencing" in search results (similar to what Amazon has done for goods it sells through its marketplace for third-parties versus sells directly from its online store (Khan, 2017). Airlines running CRSs also allowed other airlines to pay for "co-host" status to avoid some screen bias (similar to "Sponsored Ads" on Amazon and Google platforms), but they greatly increased the booking fees charged to co-hosts over time, from 25 cents to 3 dollars per booking (Locke, 1989: 221–222).

By the early 1980s, 80% of all travel agencies relied on CRSs: Sabre was used by 41% of agencies, United Airlines' Apollo system by 39%, and PARS of Trans World Airlines by 15%. Travel agencies frequently used multiple reservation systems to search more broadly for flights and prices (Locke, 1989: 221; Salpukas, 1982). For this reason, owners made their reservation systems interoperable but still gave preference in the listing order to their own flights (Minick, 2000).

Congress ordered an investigation into CRS practices in 1982, while both the Civil Aeronautics Board (CAB) and the Department of Justice (DOJ) pursued their own investigations. The DOJ decided not to file antitrust litigation, but the CAB issued new rules to curtail screen bias and discriminatory co-host pricing. Nonetheless, consumer and travel agency concerns over CRS bias persisted. In a 1988 study, the Department of Transportation noted that airlines with CRSs continued to receive a disproportionate share of bookings from the travel agents that used them (Borenstein, 1992: 64). In 1992, the House passed a bill aimed at limiting American and United's dominance over CRS, but it died in the Senate (McGinley, 1992). There were also court challenges but, since there were multiple CRSs in use, the airline owners were able to avoid antitrust interventions (Minick, 2000: 912–913). Still, the airlines seem to have felt ongoing pressure to change their operations. United spun off Apollo in 1992, and American spun off Sabre partially in 1996 and then fully in 2000 (Travel Weekly, 2000). By 1999–2000, it was also becoming clear that the Internet provided another way for online travel agencies and individual customers to make airline reservations. It also appeared to analysts that Sabre could be more valuable as an independent industry platform rather than as a service controlled by one company (McCartney, 1999). In 2021, Sabre and Apollo (renamed Travelport) remain among the main four CRSs, along with Amadeus, Shares, and some regional systems, all of which now function more like industry-level neutral platforms.

The early history of computerized airline reservations systems suggests that companies are likely to favor their own products and services without some threat of government regulation or court action. United and American not only resembled Amazon and Google in self-preferencing and screen bias, but they also had to endure multiple government investigations, along with competitor and customer lawsuits. The CRS owners also learned the difference between a company-controlled product or service and a more neutral industry platform (Cusumano *et al.*, 2019). During the 1990s, they apparently concluded it was more profitable or at least less risky (in terms of possible government intervention and ongoing scrutiny) to make their systems more open to competitors and to do so without overt bias. The industry-platform strategy also enabled competitors to "grow the pie" together—that is, to increase the travel business for everyone by making it easier for individual customers and travel agencies to book flights and buy tickets from multiple airlines.

4. Post-internet: examples of self-regulation among digital platforms

The pre-Internet historical examples raise obvious questions of interest to modern digital platforms For instance, under what conditions is self-regulation likely to occur? Should firms wait for governments to impose controls on questionable behavior and respond defensively? Or should they act aggressively in advance of potentially invasive government rules and controls? We argue that history suggests intrusive government regulations can impose serious costs and consequences on digital platforms and their markets. While the costs of government action in the Internet era have been modest so far, regulatory authorities around the world have become much more determined to reign in big tech platform power. The shadow of increasing government action should encourage more preemptive self- regulation in specific sectors, as our examples illustrate.

4.1 Preemptive self-regulation: online auctions and fraud prevention

A successful case of preemptive self-regulation involved eBay and online actions. Two years after Pierre Omidyar founded eBay in 1995, company data indicated that users had reported only 27 cases of fraud. By that point, according to a company press release, eBay had successfully completed over two million auctions, believing that "the vast majority of people will conduct trades in an honest manner" (eBay, 2020). However, eBay said it lacked the resources to monitor every auction on its platform. Rather than monitor auctions directly, it provided users tools to establish trust, like its "Feedback Forum," which allowed users to develop reputations that eBay could review (eBay, 1997; Schwartz, 1998). If users' reputations were excessively negative, then eBay would remove them from the platform.

Although fraud started out as relatively rare on eBay, it quickly became a growing problem. According to the Internet Fraud Watch, a unit within the National Consumers League, online auctions were by far the most common form of online fraud in the early years of the Internet, accounting for 30% of complaints, which tripled between 1996 and 1997 (eBay, 1997). In 1998, the rate of complaints tripled again, and fraud related to online auctions accounted for 65% of the total (Burros, 1998). These complaints concerned sellers failing to deliver items they had auctioned off, shills bidding up the price of items, and counterfeit goods. The FTC then launched an operation called SafeBid to inform state and federal agencies on how to track online auction fraud. In 1998, the FTC also brought actions against hundreds of defendants accused of committing fraud on online auction platforms (Carlton and Tam, 2000). Local authorities like the New York City Department of Consumer Affairs also began investigating fraudulent transactions on eBay (Taylor, 1999).

From 1997 to 1999, the number of complaints the FTC received about online auctions increased from 107 to ~11,000 in 1999 (Ho, 2000). In January 2000, the Justice Department released a report in January 2000 calling the Internet "the new frontier of fraud" (Lohr, 2000). State lawmakers showed some interest in regulating e-commerce but were uncertain how to go about it (Bair, 1999). Federal lawmakers also showed some interest, but the general consensus was against regulation for fear of hindering the growth of an important new industry. eBay's business model was such that the company profited from each transaction, whether it was fraudulent or not. Therefore, eliminating fraudulent transactions hurts eBay's short-term bottom line. Adopting a long-term view and despite the fact that eBay made money on every transaction, the company introduced new measures in 1999 to combat fraud in response to growing criticism. These included improvements to its Feedback Forum, a new "Verified eBay User" program, harsher penalties for shill bidding and 'deadbeat' bidders who fail to honor their commitments after winning, and easy access to insurance and escrow services (eBay, 1999). eBay also put new restrictions on items that could be put up for auction. It halted the sale of all guns and ammunition in February 1999 (Reuters, 1999). Other prohibited items included human skulls and other remains, live animals, fake IDs, and explosives (Mendoza, 1999). Over time, the rate of fraud complaints eBay received decreased from 30 per million listings to 25 per million listings in 2000 (Liedtke, 2000).

During the past two decades, eBay's governance rules and practices have continued to keep fraud and other complaints at very low levels. Equally important, eBay appears to have preemptively avoided much more intrusive action by the FTC and other governmental authorities.

This type of self-regulation is different from our other examples because it was one company that regulated itself, rather than relying on multiple companies in the same industry relying on collective action. At the time, eBay dominated online auctions, especially in the United States, and coordinating with other players was unnecessary.

4.2 Self-regulation with some government attention: policing content on social media

Getting ahead of all regulation is possible, as eBay has done. But history suggests that the most common and typical pattern is some self-regulation, especially under threat or pressure from government action—suggesting that self-regulation may actually be more of a gray area between government and company action. We can see this in social media, where platform users, even more so than we saw with traditional radio and television broadcasting and advertising, have had to manage potentially harmful, misleading, or illegal content and advertisements.

A good example is YouTube, where virtually all content under Google's ownership went unsupervised for a decade after 2006. In 2017–2018, however, YouTube faced heightened scrutiny due to reports that it was allowing violent content to seep past its YouTube Kids filter, designed to block content inappropriate for children (Woollacott, 2017). YouTube also repeatedly sparked outrage for its role in perpetuating misinformation and harassing videos in the wake of mass shootings and other national tragedies (Levin, 2017). In response to increasing negative press and public sentiment, YouTube CEO Susan Wojcicki announced in December 2017 that Google was going to hire thousands of new "moderators," expanding its total workforce to more than 10,000 people responsible for reviewing content that could violate its policies (Wojcicki, 2017). In addition, YouTube announced it would continue to develop advanced machine-learning technology to flag problematic content for removal automatically.

This attempt to self-regulate seemed at first to yield positive results. By December 2017, YouTube claimed that machine learning helped its human moderators remove nearly five times as many videos as they were previously and that algorithms now flagged 98% of videos removed for violent extremism. Wojcicki claimed that technological advances allowed the site to take down nearly 70% of violent extremist content within 8 hours of being uploaded (Wojcicki, 2017).

Self-regulation also seemed good for business because the various sides of the YouTube platform were affected by problematic contents. Some advertisers pulled their ads because they were placed alongside inappropriate videos with hate speech and extremist contents. Some high-profile brands suspended YouTube and Google ads after reports revealed they were placed alongside videos filled with sexually explicit or exploitative content about children. YouTube then announced in December 2017 a reform of its advertising policies, saying it would apply stricter criteria and conduct more manual curation as well as expand the team of ad reviewers.

In January 2018, YouTube announced that videos from its most popular channels would be subject to human review, preemptively checking large amounts of contents to ensure it met "ad-friendly guidelines" (Hern, 2018). By doing so, YouTube raised the bar for video creators who wished to run ads on their content, while hoping to allay advertiser unease about the video-sharing website. Advertisers were now able to choose ads on channels verified as "Google Preferred," which would be manually reviewed and decide which ads would only run on verified videos. YouTube announced it would complete manual reviews of Google Preferred channels and videos by March 2018 in the United States and all other markets where it offered Google Preferred.

To prevent and counter the spread of illegal hate speech online, in May 2016, the European Commission began working with Facebook, Microsoft, Twitter, and YouTube on a "Code of Conduct on countering illegal hate speech online." In 2018, Instagram, Snapchat, and Dailymotion adopted the Code of Conduct. Jeuxvideo.com joined in January 2019, and TikTok came on board in September 2020, bringing to eight the number of platforms that agreed to abide by the Code (European Commission, 2020a). Adoption of the Code includes a regular monitoring exercise setup in collaboration with a network of organizations located in the different European Union (EU) countries. Using a commonly agreed-upon methodology, these organizations test how the platform companies are implementing the commitments in the Code.

Self-regulation initially appeared to encourage the European Commission to back away from proposing binding legislation to force online platforms to remove posts containing hate speech. According to Commission data from January 2018, Twitter, YouTube, Facebook, and Microsoft reviewed about 82% of complaints about hate speech within 24 hours. This level was a significant change from May 2017, when the firms reviewed only 39%. By January 2018, Facebook had removed 79% of posts containing hate speech across the EU, while YouTube took down 75% and Twitter removed 46% (European Commission, 2018). In June 2020, the European Commission's latest assessment of the non-legally binding agreement lauded "overall positive" results—with 90% of flagged content assessed within 24 hours and 71% of content deemed illegal hate speech removed. The latter was up from just 28% in 2016. However, the European Commission also found in its fifth biannual monitoring exercise, published in June 2020 and assessing the effects of this Code of Conduct, that platforms were still lacking in transparency. Nor were they providing users with adequate feedback on the issue of hate speech removals (Lomas, 2020). While platforms gave feedback to 67% of the notifications received, only Facebook informed users systematically, with the Commission noting that "all the other platforms have to make improvements." Another criticism by the European Commission relates the inconsistencies in the platforms' evaluation processes, with assessments of flagged content that were showing "divergences" in how they were handled.

Another European Commission-led voluntary code of conduct, the Code of Practice on Disinformation, was signed on a voluntary basis in October 2018 by Facebook, Google, Twitter, Mozilla, and advertisers, with Microsoft joining in May 2019 and TikTok in June 2020 (European Commission, 2021). Signatories are asked to report monthly on their actions to improve scrutiny of ad placements, ensure transparency of political and issue-based advertising, and tackle fake accounts and malicious use of bots.

More recently, EU lawmakers have signaled a possible move away from relying on self- regulation of hate speech. They have drafted a new Digital Service Act that indicates the EU may update its rules around online liability and seek to define platform responsibilities concerning online content. The Digital Services Act represents the EU's ambitious plan to regulate online services and will cover areas of the platform economy ranging from liability for content, market dominance, and advertising to safety, smart contracts, online self-employment, and future governance frameworks. The European Commission unveiled the package of measures in December 2020 (European Commission, 2020d).

The nearing advent of the new European Digital Service Act seems to have propelled industry actors to make concrete propositions to regulators. For example, EDiMA, the Brussels-based European Digital Media Association (renamed Dot Europe in November 2020), a digital industry lobby group including Google, Facebook, Airbnb, Amazon, Apple, eBay, Microsoft, TikTok, Spotify, and Twitter, published in 2020 an "Online Responsibility Framework." This argues that online service providers should have a "responsibility to take reasonable, proportionate and feasible actions to mitigate the presence of illegal content or activity on their services." It advocates for an initial focus on illegal content rather than on "harmful" activity or content because the latter often falls into a gray zone at the EU level. It may be illegal in some member states yet considered "not-illegal-but-harmful" in others. In EDiMA's view, "harmful" content poses serious problems online but, due to the fact that it is not always illegal, members cannot act in the same way as when there is illegal content. The document adds "With content/activity that is 'harmful' but not illegal, often a service provider will have to make determinations as to where to draw the line between free speech and the right to information versus the possible harm caused to users." They provide the example of public insult of religion or blasphemy, which is not illegal in Denmark or France but is illegal in Germany, Poland, Spain, and Italy. EDiMA's position is that "mandating that content deemed 'not-illegal-but-harmful' be removed"—particularly in a pan-European law covering many different legal systems and cultural contexts—may adversely impact users' freedom of speech and expression (EDiMA, 2020). In an interview, EDiMA's director-general indicated that "It shouldn't be up to our industry to define the legal context for every EU member state, because every context is wildly different (Stolton, 2020a)."

Facebook, Google, and other platforms probably could have avoided some current difficulties if they had pursued more specific governance or self-regulation measures earlier and more

vigorously. They could have done this not only individually but more effectively if they had rallied other platforms to establish and enforce a strong code of conduct. While some platforms agreed to sign on to this government-driven code of conduct, public, and government perceptions still seem to be that the big platforms have not done enough to restrict harmful content. The digital platforms' latest proposals indicate that they now see regulation as a real possibility, so they are more seriously beginning to self-regulate and propose innovative solutions that straddle the line between free speech and the propagation of disinformation.

Some of the most decisive moves from social media platforms have been to slow down the viral relaying of false information. For example, Twitter's decision to label some posts as misleading—despite persistent criticism from President Trump and other conservatives—appears to have reduced the spread of misinformation (Wells, 2020). Twitter said those who viewed tweets to which Twitter applied a label or warning message shared them 29% less. Recent research also indicates that if a platform indicates via crowdsourcing or third-party fact-checkers that certain information is potentially biased, users will take more care in evaluating that information (Pennycook and Rand, 2019; Pennycook *et al.*, 2020). However, another study also indicated that users were likely to assume untagged information is accurate (Schwartz, 2017). This latter finding presents an enormous challenge to the social media platforms since the vast majority of their content remains untagged.

Another example of a social media platform that started to self-regulate after the threat of legislation and a public outcry was Pornhub, a video-sharing platform offering user-generated content of a graphic, sexual nature. As of December 2020, Pornhub attracted 3.5 billion visits a month, more than Netflix, Yahoo, or Amazon. Pornhub's business model was similar to YouTube: it generated revenue from advertising, while allowing users to watch videos for free. Since Pornhub's launch in 2007, any user could upload content to the site. A lucrative business, Pornhub delivered almost 3 billion ad impressions a day. One ranking listed Pornhub as the 10th-most-visited website in the world (Kristof, 2020a). Pornhub also tried to cultivate a friendly public image, combining donations to charities with support for the Black Lives Matter movement (Nathaniel, 2020). It also offered its Premium service for free during the COVID pandemic to "help flatten the curve" by encouraging people to stay home and watch "adult entertainment (Mehta, 2020)."

In 2018, the US Congress limited the shield provided by Section 230 of the US Communication Decency Act. For the first time, the new legislation allowed state and civil lawsuits against platforms for "knowingly assisting, supporting, or facilitating" online sex trafficking. In response, Pornhub started to do more self-policing by increasing its content curation through human moderation (Selyukh, 2018). Yet, apparently, Pornhub's self-regulation was ineffective. In December 2020, the *New York Times* published an in-depth investigation that followed the lives of child sexual abuse victims whose videos were uploaded to the platform. This revealed that the Pornhub website was "infested with rape videos" and "monetized child rapes, revenge pornography, spy videos of women showering, racist and misogynist content, and footage of women being asphyxiated in plastic bags" (Kristof, 2020a). While displaying videos of consenting adults was legal in the United States, displaying rape videos or videos showing adult sexual activities with children was illegal as well as widely considered to be morally abhorrent.

The reaction to the *New York Times* article sparked public outrage and swift corrective actions by Pornhub and some of its key business partners. One day after the publication, four US senators introduced bipartisan legislation making it easier for rape victims to sue porn companies that profit from videos of their assaults. There was additional bipartisan US legislation under discussion to regulate companies like Pornhub more rigorously. The Prime Minister of Canada also announced that his government was developing new regulations for these platforms, especially since Pornhub was based in Canada (Kristof, 2020b). Visa and Mastercard responded to the *New York Times* article by announcing they would stop processing credit card transactions on the website. In response, Pornhub, in a fundamental shift, decided to change the rules of who would be allowed to publish content: within a week, only verified users could publish content, and Pornhub purged almost 9 million videos (out of 13.5 million) from its platform (Cole, 2020).

Overall, the actions of social media and digital content platforms to flag misinformation or harmful content reflected a path of convergence with regulators. In September 2020, the European Commission Vice-President for Values and Transparency, Věra Jourová, indicated that the

Commission's intention was to focus on how to limit the spread of such content online: "In order to address disinformation and harmful content we should focus on how this content is distributed and shown to people rather than push for removal" (Stolton, 2020b).

4.3 Self-regulation with intense government attention: ISIS recruitment and child abuse bans

While there are many different types of platforms, ranging from video sharing (YouTube) to messaging (Twitter, WhatsApp) and social media (Facebook), the number of platforms that needed to self-regulate was relatively low. Since each sector was highly concentrated, the coordination costs were more manageable compared to fragmented industries. The critical challenge was how to coordinate without seriously impacting each firm's competitive position. While the leading players dominated their respective segments, they were indirect competitors for digital advertisers. The example of Facebook, Twitter, Microsoft, and Google YouTube coming together to police ISIS recruiting and "terrorist content" (Clark, 2016) as well as child abuse may be the strongest case for the potential of self-regulation and how to solve coordination problems.

The stimulus for regulating terrorist content came in 2014 when the UK's Intelligence and Security Committee (ISC) reported that one of the killers of Fusilier Lee Rigby had written previously about his intent to kill a soldier on Facebook. The ISC accused Facebook and other platforms of providing a "safe haven" for terrorists and demanded they do more to inform law enforcement of potential terrorist threats (Dodd *et al.*, 2014). Then, in late 2015 and early 2016, Europe suffered two major terrorist attacks, one in Paris and another in Brussels. There was also a major terrorist attack in San Bernardino, California, in December 2015. Shortly after the Brussels attack, the Council of the EU released a statement promising to "intensify work with IT companies … to counter terrorist propaganda" (European Council, 2016).

The motivation for self-regulation, in this case, was clearly the threat from European lawmakers to introduce legislation that would compel them to police their platforms EU Justice Commissioner Věra Jourová explicitly warned companies that the EU was considering legislation to address the problem (Clark, 2016). The initial response by the largest American platforms to this threat was to offer in May 2016 a voluntary Code of Conduct on Countering Illegal Hate Speech Online. The Code asserted that they "support the European Commission and EU Member States in the effort to respond to the challenge of ensuring that online platforms do not offer opportunities for illegal online hate speech to spread virally." One of the commitments the Code sets for them is to "intensify cooperation between themselves and other platforms and social media companies to enhance best practice sharing" (European Commission, 2020c). The Global Internet Forum to Counter Terrorism (GIFCT) was also established in 2017 to foster this cooperation and, by 2021, it had expanded to 13 members (https://gifct.org/about/).

Prior to the establishment of the GIFCT, Facebook, Twitter, YouTube, and Microsoft committed to creating a shared database of "hashes"—unique digital fingerprints—of terrorist propaganda videos and images that would enable them to identify such content the moment it was uploaded and prevent the content from ever being posted or shared. The GIFCT manages the database, which is accessible to members of the Hash Sharing Consortium. Because companies have different definitions of terrorism or terrorist content, the Consortium originally focused on "content relating to groups appearing on the United Nation Security Council's consolidated sanctions lists" (GIFCT, 2020: 3). A more specific taxonomy of content was subsequently established. The hashes in the Consortium's database represent mostly instances of "Glorification of Terrorist Acts" (72% of hashes) and "Graphic Violence Against Defenseless People" (17%). The third most common hash represents content falling under the GIFCT's Content Incident Protocol (CIP), which concerns content created by a perpetrator during the course of a real-world attack. Some 9% of hashes in the database fall under the CIP and most of these hashes represent content produced by the Christchurch shooter who streamed his attacks in 2019. The remainder of the hashes in the database included content aimed to recruit, radicalize, or instruct others (2%) and imminent credible threats (0.1%). In total, there are 300,000 unique hashes in the database representing 250,000 images and 50,000 videos.

Social media platforms gradually changed their approach to terrorist content to answer the demands of European lawmakers responding to terrorist attacks. When a member finds content

on its platform that meets the criteria of the Consortium, it creates a hash to represent the content and adds it to the database. Other companies can add labels that indicate they think the hash has been categorized incorrectly or to dispute that it is terrorist content. GIFCT enabled members to add these labels "so that third companies can make their own decision on how best to use the hashes … dependent on their own processes and review systems" (GIFCT, 2020: 5).

In addition to sharing hashes, members of the GIFCT also share URLs that link to other members' platforms—for example, a link on Facebook to a post on Twitter. GIFCT members share links to give other platforms the chance to review the content and decide whether to remove it. The GIFCT has partnered with SITE Intelligence, a firm that monitors and analyzes terrorist content, and runs a dashboard that provides context around URLs, like the affiliation of the terrorist content and translations of the content. Through its program, the GIFCT has shared 24,000 URLs (https://gifct.org/about/story/).

The GIFCT also administered a real-time response process called the CIP. The purpose of the CIP was to respond quickly to events like the Christchurch shooting, which was streamed live. The CIP involved sharing hashes of the content with GIFCT members and establishing a stream of communication between all members to quickly respond to developments. It was designed "to be easily integrated into external crisis response procedures," like the EU's Crisis Response Protocol, which the EU described as a "voluntary mechanism to help coordinate a rapid, collective and cross-border response to the viral spread of terrorist and violent extremist content online" (GIFTC, 2019).

Representatives of GIFCT members tested the CIP with European law enforcement authorities during a tabletop exercise held at The Hague in September 2019 (Microsoft Corp, 2017). In 2019, the GIFCT released its first transparency report, which acknowledged that it had added more than 200,000 hashes of terrorist content to its shared database. A few months later, it was credited with preventing the spread of footage of an anti-Semitic terrorist attack in Germany, which had inspired the shootings in New Zealand (Uberti, 2019).

There have been criticisms and concerns about GIFCT. Its definitions of what constitutes "terrorist content" are somewhat vague and decisions have lacked transparency. It has also made operational mistakes, such as removing anti-terrorist advocacy groups content. It could have little or no effect if producers of terrorist content simply move to other platforms. In July 2020, Human Rights Watch and 14 other human rights groups signed a letter to GIFCT expressing concern over the lack of visibility to outsiders into what content, exactly, is included in GIFCT's shared hash database (Human Rights Watch, 2020).

Simply belonging to this coalition did not mean that all members had similar results in tackling content, but it had a positive impact on member activities. For example, in the first three months of 2020, Facebook took action on 6.3 million pieces of content violating its rules against terrorist content (Facebook, 2020). More than 99% of the terrorist content Facebook removed was caught by proactive detection technology. By comparison, Twitter used to remove around 40% of violent or terrorism-related content proactively; in 2020, it removed 74% proactively (U.S. Senate Committee on Commerce, Science, & Transportation, 2019; Twitter, 2019). Due to Facebook's changes to its detection technology, the pieces of content Facebook took action on increased in the second quarter of 2020 to 8.7 million (Facebook, 2020).

The GIFCT was not the first instance of tech companies collaborating by sharing data or technology. In fact, the method that the GIFCT used to catalog terrorist content was based on an earlier method called PhotoDNA, developed by Microsoft to track and quickly remove child sexual abuse material (CSAM), which has been explicitly illegal in most countries. PhotoDNA went live on skyDrive (now Microsoft OneDrive) and Bing in 2009 and was later implemented on Facebook, Twitter, and Google's various platforms between 2010 and 2016. However, while CSAM would seem to be an obvious candidate for self-regulation, tech companies were initially hesitant to act. At the time, identifying CSAM was considered a difficult computational problem and prescreening content before it was posted would drastically slow down platform operations. In addition, screening content in this way presented privacy issues, depending on how it was deployed. Improvements in processing power plus an enormous database of known CSAM from the National Center for Missing & Exploited Children (NCMEC) made building a tool for identifying CSAM more straightforward (Farid, 2017).

Still, unlike hate speech and even terrorist content, identifying CSAM was generally less controversial, and there was more public tolerance for false positives. In addition, Microsoft and the companies that implemented PhotoDNA feared further regulation that would require them to actively screen for CSAM and impose sanctions if they did so ineffectively. Finally, the incentives for collaboration were higher than many other areas because the NCMEC's database of CSAM was not in the hands of any one company. Legal scholar Evelyn Douek argued the following:

> [A]s public pressure and controversy around content decisions increases and becomes increasingly costly, the pressure not to compete on things not core to platforms' business models may also increase … Content moderation is hard, and presenting a united front can prevent any individual company from receiving too much opprobrium for any particular decision…. Particularly in difficult or controversial areas, it could become increasingly attractive for platforms to adopt a kind of default position instead of facilitating the kind of diversity that could help continually reevaluate the bounds of accepted expression (Douek, 2020).

Why and when will companies collaborate on issues like screening for CSAM or terrorist content? Part of the answer seems to be that, with the threat of restrictive legislation, each firm viewed the reputational risks of doing nothing as outweighing the immediate downsides of doing something. Relatedly, companies seem to view the reputational benefits of collaboration as outweighing the upside of any potential competitive advantages (e.g., having the resources to develop better CSAM screening technology than your competitor).

One important difference between moderating terrorism and CSAM versus moderating hate speech and misinformation was the intensity of regulatory threats and the degree of consensus on the problem definition and solution. There was virtually universal consensus on banning terrorism and sexual exploitation, but hate speech was far more controversial, with little consensus on basic definitions.

5. Analysis and discussion

We have reviewed several examples of regulatory initiatives and attempts at self-regulation from before and after the Internet. These cases suggest a pattern. All our examples, from before and after the Internet era, involved cases when there was a gap or vacuum between government regulation versus platform self-regulation and ecosystem governance. This vacuum probably accompanies the birth of many new industries that do not start out with strict government oversight. At least in the cases we examined, after the first few years of a "wild west" environment, governments stepped in to regulate or pressure firms to curb abuses. This pattern occurred with movies, radio, and television broadcasting and advertising and airline reservations via computers. In the Internet era, to avoid invasive government regulation, digital platforms have been introducing some controls on behavior and usage. Technology that exploits big data, artificial intelligence, and machine learning, with some human curation, will increasingly give digital platforms the ability to monitor and regulate more of what happens on their platforms. The unresolved question is to what extent leaders of the dominant digital platforms have the *will* to continue their efforts to self-regulate and govern their platforms and ecosystems more tightly, even if these efforts harm revenues and profits in the short term. The examples also suggest several insights for managers of digital platforms and government regulators.

First, the argument for more self-regulation by digital platforms rests on the recognition that these companies have extraordinary economic, social, and political power that will continue to grow. Platform managers must learn when to use and when not to use their considerable influence. This means they must figure out how to compete aggressively while staying within the bounds of what different societies consider legal as well as fair and ethical. To avoid the moral hazard or tragedy of the commons dilemma that we noted earlier, managers and boards of directors need to figure out how not to destroy public trust, which could harm the market (both demand and supply sides) for all digital platforms. Digital platform managers also need to anticipate when government regulation is likely to become a threat to their business models.

Second, we find that firms and new industries tend not to pursue self-regulation when the perceived short-term costs are high—when changes imply a depressing of network effects and a potential reduction in revenue or profit. This is to be expected; no firm likes industry regulations that are "bad for business," as we have seen with movie and video-game ratings as well as traditional and online advertising. This is a problem for social media platforms in particular because fake news stories and damaging videos, and reports of spectacular conspiracy theories, are more frequently read and forwarded than real news items, and they generate more activity, stronger network effects, and more advertising revenues (Aral, 2020). However, the history of movies, video games, and other sectors suggest that avoiding self-regulation can also harm the business in the long term. If governments impose intrusive regulations, or if platform misuse and abuse undermines consumer trust and demand, then the business of digital platforms may well decline.

Third, firms are more likely to get serious about self-regulation when they see a credible threat of government regulation, even if self-regulation may hurt short-term sales and profits. This pattern occurred with tobacco and cigarette ads, airline reservations, and social media ads for terrorist recruitment as well as pornographic material. We also see this playing out when companies face government licensing regulations. Licensing gives a government more clout and more opportunities to intervene in a market or to prevent firms from operating, as in the movie industry or broadcasting over the public airwaves. The historical pattern reinforces our suggestion that firms should not wait too long to self-regulate, lest government agencies impose standards, rules, or licensing requirements that may be detrimental to industry growth and innovation. We also find that certain types of problems trigger self- regulation more than others, such as when the threat of government invention is related to fraud, health, or public safety. We saw this in the limits for advertising of alcohol and tobacco products and restrictions on terrorist group recruitment and sexual exploitation via online content. Under these circumstances, governments have greater incentives to intervene, which means that firms have greater incentives to preempt aggressive regulation.

Fourth, self-regulation need not occur only at the level of the firm. Sometimes coalitions of firms in the same market, with presumably similar business models and economic incentives, have worked together effectively to create common norms and standards. We saw this cooperation in movie and video- game rating systems, television advertisements curbing unhealthy products like alcohol and tobacco, computerized online airline reservations that stopped giving preference to the system owners, and codes of conduct as well as bans or tags on terrorist advertisements and misinformation in social media. Individual firms may hesitate to enact self-regulation if they incur added costs that their competitors do not. This is why industry-wide coalitions have the benefit of reducing free riding. Such an argument is consistent with the "Enforceable code of conduct," proposed by the U.K. government (U.K. Competition and Markets Authority, 2020) and the European Commission (U.K. Competition and Markets Authority, 2020; European Commission, 2020b). The U.K.'s CMA even recommended that a small number of the largest firms (platforms with "Strategic Market Status") should identify "high-level principles rather than detailed and prescriptive rules." The objective of a code of conduct is to create fair trading, open choices, trust, and transparency.

Fifth, regulation and self-regulation can impact platform sides differently, and the relative alignment of incentives across platform sides is very likely to impact the probability (and profitability) of self-regulation. Any regulation may be more costly for or, alternatively, be more valued by, one side compared to another. Platforms may also be tempted to favor the paying side rather than the subsidized side. For example, social media businesses may decide to curate more on the supply side (i.e., the content contributors) and less on the demand side (i.e., platform users). At the same time, platform companies can also think more positively about regulation and self-regulation. As we have seen with movies and video games, it is possible that regulation or self-regulation on the supply (content) side could create new market segments and demand, such as for higher quality content, like family-oriented or trustworthy social media. Higher quality and market segmentation may eventually contribute to expansion rather than contraction of demand and supply.

Sixth, when firms or industries impose guidelines that may violate guarantees of free speech, as with movie ratings or tags on social media, these self-regulations tend to attract internal as well as external resistance. Some resistance may be due to the complex tradeoffs associated with what different stakeholders consider acceptable and the desire of firms to increase their revenues and profits. We may also argue that it is not the role of firms to be the arbiters of free speech. However, it is clear that digital platforms can play a more active role in restricting harmful or morally questionable content. Social media companies, in particular, seem well-suited to curate content because relaying and amplifying information (as opposed to generating information) is a core activity and competence (Aral, 2020).

Seventh, the failure to self-regulate can put the benefits of limited existing regulation at risk. For example, Section 230 of the United States 1996 Communications Decency Act provides broad immunity for online intermediaries from liability for user-generated content posted on their sites. However, the lack of effective self-regulation by Facebook, Twitter, and other platforms led Donald Trump, Joseph Biden, and numerous other US politicians on both sides of the aisle to call for the elimination of Section 230 (Kelly, 2020; Edelman, 2020). In response, company CEOs seem eager to participate in any regulatory changes involving this legislation. More aggressive self-regulation and cooperation with the government might well produce a better outcome for social media platforms compared to leaving their fate solely up to the politicians. This suggests that in the future we may see a less clear demarcation between government regulation and platform self-regulation.

The puzzle remains why, in some settings, self-regulation has worked reasonably well while, in other settings, it has not worked well at all. Figure 1 summarizes where we think the "zone of self-regulation" exists, based on our review of several historical and current examples. In the absence of a threat of government regulation (bottom left and right), most firms will see little incentive to regulate themselves. Similarly, when the perceived costs of intrusive regulation are very high, most firms would rather fight (such as through lobbying or litigation) than self-regulate (upper right quadrant). In other words, for self-regulation to work well, firms should perceive the costs to be manageable *and* the threat of government regulation to be credible. As seen in the left side of the graph, when the perceived costs of self-regulation are not excessive, and there is

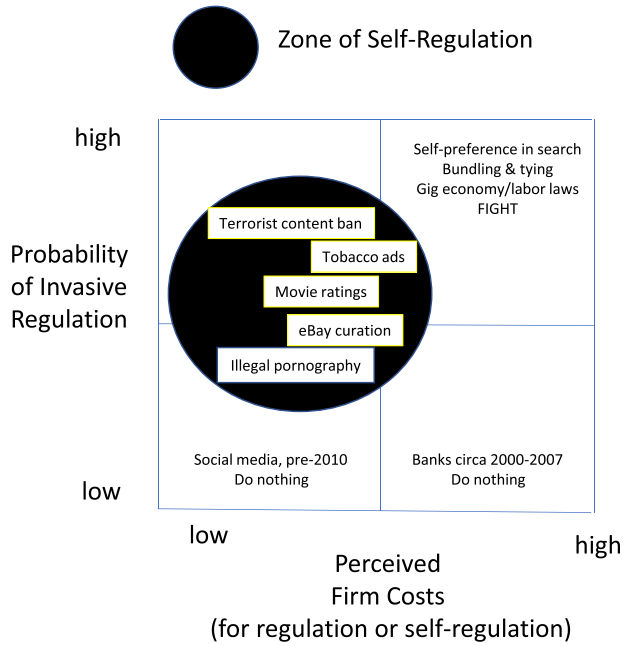

Figure 1. Zone of Self-Regulation.

a growing threat of government intervention, firms and industries should have greater incentives to self-regulate and get ahead of the regulation curve. Social media's ban of terrorist content and illegal pornographic material, eBay's effort to limit fraud, and movie ratings, as well as curbs on tobacco ads, fit this description.

6. Concluding comments

In this article, we tried to clarify when firms are likely to attempt self-regulation, why they sometimes choose either not to do so early on, or do so without much enthusiasm. We worry that short-term thinking may prevent business leaders from seeing the benefits of self-regulation, such as higher quality content for users and healthier industry growth. We also identified some areas likely to be better handled by self-regulation and platform governance rules, such as quality, health, or safety, especially when the technology or platform operations are difficult for government regulators to understand and monitor. However, neither pure self-regulation nor aggressive government regulation seems likely to cover all the challenges digital platforms face.

Other authors, including in this special issue (Kwoka and Valletti, 2021), agree with us that most firms will not self-regulate without government pressure. However, most governments do not have the skills or resources to regulate and monitor the dynamic, ongoing changes inevitable with digital platforms and their complex technologies and operations. Yet, to avoid cases of moral hazard or even a broader tragedy of the commons, where internet markets are diminished by a lack of trust or intrusive regulation, governments and firms will need to work together. This dilemma leads to important questions for further research: how can digital platforms and governments collaborate to create better regulatory outcomes for their unique businesses? How can governments incentivize powerful firms to "do the right thing?" Since more aggressive government oversight over Google, Facebook, Twitter, Amazon, and other digital platforms seems inevitable, new institutional mechanisms for more participative forms of regulation would seem to be critical to their long-term survival and success.

Funding

Michael Cusumano's research was supported by the MIT Sloan School of Management. Annabelle Gawer's research was supported by Surrey Business School, University of Surrey. David Yoffie's research was supported by the Harvard Business School Division of Research. Harvard Business School research associate Daniel Fisher assisted in the research.

References

Aral, S. (2020), *The Hype Machine: How Social Media Disrupts Our Elections, Our Economy, and Our Health – and How We Must Adapt*. Penguin Random House: New York.

Arrow, K. (1963), 'Uncertainty and the welfare economics of medical care,' *The American Economic Review*, **53**(5), 941–973.

Australian Competition and Consumer Commission (ACCC). (2019), 'Digital Platforms Inquiry: Final Report,' https://www.accc.gov.au/publications/digital-platforms-inquiry-final-report.

Bailey, J. (2011), '10 films that avoided the NC-17 and suffered for it,' *The Atlantic*, December 2.

Bair, J. (1999), 'Lawmakers get briefing on electronic commerce as they consider regulations,' *Associated Press*, September 30.

Baldwin, C. and J. Woodard (2009), 'The architecture of platforms: a unified view,' in A. Gawer (ed.) *Platforms, Markets and Innovation*. Edward Elgar: Cheltenham, pp. 19–44.

Borenstein, S. (1992), 'The evolution of U.S. Airline competition,' *Journal of Economic Perspectives*, **6**(2), 45–73.

Boudreau, K. and A. Hagiu (2009), 'Platform rules: multi-sided platforms as regulators,' in A. Gawer (ed.) *Platforms, Markets and Innovation*. Edward Elgar: Cheltenham, pp. 163–191.

Burns, D. *et al.* (1997), 'Cigarette smoking behavior in the united states,' Monograph 8: changes in Cigarette-Related Disease Risks and Their Implications for Prevention and Control. National Cancer Institute, pp. 13–16, https://cancercontrol.cancer.gov/brp/tcrb/monographs/.

Burros, M. (1998), 'Lots of antiques on line, but not many rules,' *New York Times*, November 12.

Campbell, A. J. (1999), 'Self-regulation and the media,' *Federal Communications Law Journal*, **51**(3), 711–771.
Carlton, J. and P. Tam (2000), 'Online auctioneers face growing fraud problem,' *Wall Street Journal*, May 12.
Clark, L. (2016), 'Facebook and Twitter must tackle hate speech or face new laws,' *Wired*, December 5.
Coglianese, C. and E. Mendelson (2010), 'Meta-regulation and self-regulation,' in M. Cave, R. Baldwin and M. Lodge (eds), *The Oxford Handbook on Regulation*. Oxford University Press: Oxford, pp. 146–168.
Cole, S. (2020), 'Pornhub just purged all unverified content from the platform,' *Motherboard, Tech by Vice*. December 14.
Conant, M. (1978), *Antitrust in the Motion Picture Industry*. Arno Press: New York, p. 40.
Corcoran, E. (1994), 'Video, computer game industries split on ratings,' *Washington Post*, July 29, 1994, accessed via Factiva, August 12, 2020.
Couvares, F. G. (1992), 'Hollywood, main street, and the church: trying to censor the movies before the production code,' *American Quarterly*, **44**(4), 592–593.
Cummings, K. M. and R. N. Proctor (2014), 'The changing public image of smoking in the United States: 1964-2014,' *Cancer Epidemiology, Biomarkers & Prevention*, **23**(1), 32–36.
Cusumano, M., A. Gawer and D. Yoffie (2019), *The Business of Platforms: Strategy in the Age of Digital Competition, Innovation, and Power*. Harper Business: New York.
Dodd, V., E. MacAskill and P. Wintour (2014), 'Lee Rigby murder: Facebook could have picked up killer's message – report,' *Guardian*, November 26.
Douek, E. (2020), 'The rise of content cartels,' *Knight First Amendment Institute*, February 11. https://knightcolumbia.org/content/the-rise-of-content-cartels.
eBay. (1997), 'Internet Fraud Virtually Nonexistent at eBay.com; Pioneer of person-to-person online auctions safeguards against fraud with self-policing trading environment,' *Business Wire*, September 18,
eBay. (1999), 'eBay launches the most comprehensive trust and safety upgrades to the world's largest person-to-person trading site,' *PR Newswire*, January 15.
eBay. (2020), 'Internet Fraud Virtually Nonexistent at eBay.com; Pioneer of person-to-person online auctions safeguards against fraud with self-policing trading environment,' *Business Wire*, September 18, 1997, via Factiva, accessed August 26, 2020.
Edelman, G. (2020), 'On Section 230, it's trump vs. trump,' *Wired*, December 3.
EdiMA. (2020), 'Fundamentals of the online responsibility framework series: a legal basis to act,' https://www.euractiv.com/wp-content/uploads/sites/2/2020/10/ORF-Series_-Basis-to-Act_EDiMA.pdf.
Eisenmann, T., G. Parker and M. Van Alstyne (2006), 'Strategies for two-sided markets,' *Harvard Business Review*, October, 92–101.
eMarketer. (2020), 'Facebook ranks last in digital trust among users,' *eMarketer.com*, September 24.
Entertainment Software Rating Board. 2021), 'Frequently asked questions,' https://www.esrb.org/faqs/#who-decides-which-rating-should-be-assigned, July 10, 2021.
European Commission. (2018), 'Results of Commission's Last Round of Monitoring of the Code of Conduct Against Online Hate Speech,' January http://ec.europa.eu/newsroom/just/item-detail.cfm?item_id=612086, October 30, 2018.
European Commission. (2020a), 'The Code of Conduct on Countering Illegal Hate Speech Online. https://ec.europa.eu/commission/presscorner/detail/en/qanda_20_1135.
European Commission. (2020b), 'Digital Markets Act,' https://ec.europa.eu/info/sites/info/files/proposal-regulation-single-market-digital-services-digital-services-act_en.pdf.
European Commission. (2020c), 'The EU Code of Conduct on Countering Illegal Hate Speech Online,' https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en.
European Commission. (2020d), 'Europe Fit for the Digital Age: Commission Proposes New Rules for Digital Platforms,' December 15. https://ec.europa.eu/commission/presscorner/detail/en/ip_20_2347.
European Commission. (2021), 'Code of Practice on Disinformation,' https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation.
European Council. (2016), 'Joint statement of EU Ministers for Justice and Home Affairs and representatives of EU institutions on the terrorist attacks in Brussels on 22 March 2016,' March 24, https://www.consilium.europa.eu/en/press/press-releases/2016/03/24, accessed September 24, 2020.
Facebook. (2020), 'Community Standards Enforcement Report,' https://transparency.facebook.com/community-standards-enforcement#dangerous-organizations.
Farid, H. (2017), 'Reining in Online Abuses,' September 5, University of California at Berkeley, video. https://www.ischool.berkeley.edu/files/reining-online-abuses-hany-farid.
Federal Aviation Authority. (2017), 'A Brief History of the FAA,' https://www.faa.gov/about/history/brief_history/, accessed July 10, 2021.

Federal Trade Commission. (2009), 'A Report to Congress,' https://www.ftc.gov/sites/default/files/documents/reports/marketing-violent-entertainment-children-sixth-follow-review-industry-practices-motion-picture-music/p994511violententertainment.pdf, December 12, 2020.
Freeman, J. (2000), 'Private parties, public functions and the new administrative law,' *Administrative Law Review*, **52**, 813–858.
Furman, J., *et al*. (2019), 'Unlocking digital competition: report of the digital competition expert panel,' UK government publication, HM Treasury.
Gardner, M. N. and A. M. Brandt (2006), 'The doctors' choice is America's Choice': the physician in US cigarette advertisements, 1930-1953,' *American Journal of Public Health*, **96**(2), 222.
Gawer, A. (2014), 'Bridging differing perspectives on technological platforms: toward an integrative framework,' *Research Policy*, **43**(7), 1239–1249.
Gawer, A. (2020), 'Digital platforms' boundaries: the interplay of firm scope, platform sides, and digital interfaces,' *Long Range Planning*, 102045.
Gawer, A. and M. Cusumano (2002), *Platform Leadership: How Intel, Microsoft and Cisco Drive Industry Innovation*. Harvard Business School Press: Boston.
Gawer, A. and N. Srnicek (2021), 'Online platforms: economic and societal effects,' *Panel for the Future of Science and Technology*, European Parliament, Study 10-3-2021.
GIFCT. (2020), 'GIFCT Transparency Report July 2020,' https://gifct.org/wp-content/uploads/2020/10/GIFCT-Transparency-Report-July-2020-Final.pdf.
GIFCT. (2019), 'A Europe that Protects – EU Crisis Protocol: Responding to Terrorist Content Online,' October https://ec.europa.eu/home-affairs/sites/default/files/what-we-do/policies/european-agenda-security/20191007_agenda-security-factsheet-eu-crisis-protocol_en.pdf.
Gould, J. (1962), '2 networks at odds with collins on cigarette ads,' *New York Times*, November 29.
Gunningham, N. (1995), 'Environment, self-regulation, and the chemical industry: assessing responsible care,' *Law and Policy*, **17**(1), 57–109.
Gunningham, N., R. A. Kagan and D. Thornton (2003), *Shades of Green: Business, Regulation, and Environment*. Stanford University Press: Redwood City.
Gunningham, N. and J. Rees (1997), 'Industry self-regulation: an institutional perspective,' *Law and Policy*, **19**(4), 363–414.
Hagiu, A. and J. Wright (2015), 'Multi-sided platforms,' *International Journal of Industrial Organization*, **43**, 162–174.
Hardin, G. (1968), 'The tragedy of the commons,' *Science*, **162**, 1243–1248.
Head, R. V. (2002), 'Getting sabre off the ground,' *IEEE Annals of the History of Computing*, October-December: pp. 32–38.
Hern, A. (2018), 'YouTube to manually review popular videos before placing ads,' *The Guardian*, January 17.
Ho, D. (2000), 'Law enforcement targeting surge in online auction scams,' *Associated Press*, February 15.
Human Rights Watch. (2020), 'Joint Letter to New Executive Director, Global Internet Forum to Counter Terrorism,' July 30, https://www.hrw.org/news/2020/07/30/joint-letter-new-executive-director-global-Internet-forumcounter-terrorism#_ftn2.
Hunt, K. (2018), 'The end of American film censorship,' *JSTOR Daily*, February 28.
Jacobides, M., T. Knudsen and M. Augier (2006), 'Benefiting from innovation: value creation, value appropriation and the role of industry architectures,' *Research Policy*, **35**, 1200–1221.
Jacobides, M. and I. Lianos (2021), 'Ecosystems and competition law in theory and practice,' *Industrial & Corporate Change*, Special Issue on Regulating Platforms & Ecosystems.
Jansen, S. and M. Cusumano (2013), 'Defining software ecosystems: a survey of software platforms and business network governance,' in S. Jansen, S. Brinkemper and M. Cusumano (eds), *Software Ecosystems: Analyzing and Managing Business Networks in the Software Industry*. Edward Elgar: Cheltanham, UK, pp. 13–28.
Jaramillo, D. L. (2018), *The Television Code: Regulating the Screen to Safeguard the Industry*. University of Texas Press: Austin, p. 27.
Jenny, F. (2021), 'Competition law enforcement and regulation for digital platforms and ecosystems: understanding the issues, facing the challenges and moving forward,' *Industrial & Corporate Change*, Special Issue on Regulating Platforms & Ecosystems.
Jha, P. (2020), 'The hazards of smoking and the benefits of cessation: a critical summation of the epidemiological evidence in high-income countries,' *eLife*, March 20.
Joseph Burstyn, Inc. v. *Wilson*. (1952), 343 U.S. 495.
Kelly, M. (2020), 'Joe Biden wants to revoke Section 230,' *The Verge*, January 17.
Khan, L. M. (2017), 'Amazon's antitrust paradox,' *The Yale Law Journal*, **126**(3), 710–805.
King, A. A. and M. Lenox (2000), 'Industry self-regulation without sanctions: the chemical industry's responsible care program,' *Academy of Management Journal*, **43**, 698–716.

Kira, B., V. Sinha and S. Srinivasan (2021), 'Regulating digital ecosystems: bridging the gap between competition policy and data protection,' *Industrial & Corporate Change*, Special Issue on Regulating Platforms & Ecosystems.
Kohler, C. (2009), 'July 29, 1994: videogame makers propose ratings board to congress,' *Wired*, July 29.
Kristof, N. (2020a), 'The children of Pornhub,' *New York Times*, December 4.
Kristof, N. (2020b), 'An uplifting update on the terrible world of Pornhub,' *New York Times*, December 9.
Kwoka, J. and T. Valletti (2021), 'Scrambled eggs and paralyzed policy: breaking up consummated mergers and dominant firms,' *Industrial & Corporate Change*, Special Issue on Regulating Platforms & Ecosystems.
Laczniak, R. N., L. Carlson, D. Walker and E. D. Brocato (2017), 'Parental restrictive mediation and children's violent video game play: the effectiveness of the Entertainment Software Rating Board (ESRB) Rating System,' *Journal of Public Policy and Marketing*, **36**(1), 70–78.
Lampe, R. and S. McRae (2014), 'This Paper is Not Yet Rated — Self-Regulation in the U.S. Motion Picture Industry,' unpublished working paper https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=IIOC2014&paper_id=558.
Levin, S. (2017), 'Google to hire thousands of moderators after outcry over Youtube abuse videos,' *The Guardian*, December 5.
Liedtke, M., (2000), 'Should eBay be responsible for the behavior of its buyers, sellers?' *Associated Press*, June 25.
Locke, L. G. (1989), 'Flying the unfriendly skies,' *Harvard Journal of Law & Technology*, **2**(1), 219–237.
Lohr, S. (2000), 'Policing the internet: anyone but government,' *New York Times*, February 20.
Lomas, N. (2020), 'On illegal hate speech, EU lawmakers eye binding transparency for platforms,' *Techcrunch*, 23 June. https://techcrunch.com/2020/06/23/on-illegal-hate-speech-eu-lawmakers-eye-binding-transparency-for-platforms/.
Lupton, J. A. (2018), 'Illinois Supreme Court History: Movies and Censorship,' http://illinoiscourts.gov/Media/enews/2018/072318_censorship.asp, August 10, 2020.
Maitland, I. (1985), 'The limits of self-regulation,' *California Management Review*, **27**(3), 135.
Marsden, P. and R. Podszun (2020), *Restoring Balance to Digital Competition – Sensible Rules, Effective Enforcement*. Konrad Adenauer Stiftung: Berlin.
McCabe, D. (2020), 'Tech companies shift their posture on a legal shield, wary of being left behind,' *New York Times*, December 15.
McCartney, S. (1999), 'AMR announces its plans to spin off sabre holdings,' *Wall Street Journal*, December 14.
McGinley, L. (1992), 'Airlines face weaker control of reservations,' *Wall Street Journal*, August 13.
Mehta, I. (2020), 'Pornhub's Premium goes free worldwide to cure your lockdown blues,' *The Next Web*, March 25.
Mendoza, M. (1999), 'Bizarre and at times illegal items on online auctions,' *Associated Press*, September 9.
Microsoft Corp. (2017), 'Facebook, Microsoft, Twitter and YouTube announce formation of the Global Internet Forum to Counter Terrorism,' June 26, https://blogs.microsoft.com/on-the-issues/2017/06/26/facebook-microsoft-twitter-youtube-announce-formation-global-Internet-forum-counter-terrorism/.
Minick, A. (2000), 'Computer reservations systems, airlines, and the internet,' *Journal of Air Law and Commerce*, **65**(4), 894–926.
Mozur, P., *et al*. (2021), 'A global tipping point for reining big tech has arrived,' *New York Times*, April 20.
Mutual Film Corporation v. Industrial Commission of Ohio. (1915), 236 U.S. 230.
Nash Information Services. (2021), 'MPAA Ratings,' https://www.the-numbers.com/market/mpaa-ratings, July 10, 2021.
Nash, J. and J. Ehrenfeld (1997), 'Codes of environmental management practice: assessing their potential as a tool for change,' *Annual Review of Energy and the Environment*, **22**, 487–535.
Nathaniel, A. (2020), 'Pornhub Stands in Solidarity Against Racism and Social Injustice,' Official Pornhub tweet on Twitter. https://twitter.com/pornhub/status/1266929094329016325?lang=en.
National Association of Radio and Television Broadcasters. (1952), 'The Television Code,' https://worldradiohistory.com/Archive-NAB-Publications/NAB-The-Television-Code-1952.pdf.
National Association of Radio and Television Broadcasters. (1954), 'Television Code Gets Recognition, NARTB Reports,' *Variety*, March 3.
Olson, M. (1965), *The Logic of Collective Action*. Harvard University Press: Cambridge.
Parker, G., G. Petropoulos and M. Van Alstyne (2021), 'Platform Mergers and Antitrust,' *Industrial & Corporate Change*, Special Issue on Regulating Platforms & Ecosystems.
Parker, G., M. Van Alstyne and S. Choudary (2016), *Platform Revolution: How Networked Markets are Transforming the Economy and How to Make Them Work for You*, WW Norton & Company: New York.
Pennycook, G. *et al*. (2020), 'The implied truth effect: attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings,' *Management Science*. online, published in advance, 4944–4957.

Pennycook, G. and D. G. Rand (2019), 'Fighting misinformation on social media using crowdsourced judgments of news source quality,' *Proceedings of the National Academy of Sciences*, **116**(7), 2521–2526.
Petit, N. and D. J. Teece (2021), 'Big tech, big data, and competition policy: favoring dynamic over static competition,' March 22. https://ssrn.com/abstract=3229180.
Piepenburg, C. (2018), 'Not Yet rated: self-regulation and censorship issues in the U.S. Film industry,' *UCLA Entertainment Law Review*, **25**(1), 98–131.
Pollay, R. W. (1994), 'Promises, promises: self-regulation of US cigarette broadcast advertising in the 1960s,' *Tobacco Control*, **3**(2), 138.
Pondillo, R. (2010), *America's First Network TV Censor*, Southern Illinois University Press: Carbondale.
PR Newswire. (1999), 'PA State Sen. Wagner introduces legislation addressing video game violence,' *PR Newswire*, April 28.
Puga, A. (1993), 'Violence in video games prompts Senate grilling,' *The Boston Globe*, December 10.
Rees, J. (1994), *Hostages of Each Other: The Transformation of Nuclear Safety since Three Mile Island*, University of Chicago Press: Chicago.
Rees, J. (1997), 'The development of communitarian regulation in the chemical Industry,' *Law and Policy*, **19**, 477.
Reuters. (1999), 'Ebay bans firearms auctions,' *Wired.com*, February 22.
Rochet, J. and J. Tirole (2003), 'Platform competition in two-sided markets,' *Journal of the European Economic Association*, **1**(4), 990–1029.
Russ, E. K. (1982), 'Overview of INPO and its programs [institute of nuclear power operations],' *U.S Department of Energy Office of Scientific and Technical Information*, **23**, 6.
Salpukas, A. (1982), 'Airlines tussle with agents,' *New York Times*, July 6.
Schwartz, E. I. (1998), 'At on-line auctions, good, and raw, deals,' March 5, accessed via Factiva, August 26, 2020.
Schwartz, J. (2017), 'Tagging fake news on Facebook doesn't work, study says,' *Politico*, September 11.
Scott Morton, F. *et al*. (2019), 'The stigler report: committee for the study of digital platforms, market structure and antitrust subcommittee, University of Chicago booth school of business, stigler center for the study of the economy and the State.'
Segerson, K. and T. J. Miceli (1999), 'Voluntary approaches to environmental protection: the role of legislative threats,' in C. Carraro and F. Leveque (eds), *Voluntary Approaches in Environmental Policy*. Springer: Dordrecht, pp. 105–120.
Selyukh, A. (2018), 'Section 230: a key legal shield for Facebook, Google is about to change'. *All Things Considered, NPR*.
Sezey, R. D. (1952), 'Give the television code a chance,' *The Quarterly of Film Radio and Television*, **7**(1), 5.
Sinclair, D. (1997), 'Self-regulation versus command and control? Beyond false dichotomies,' *Law & Policy*, **19**, 529–559.
Sorrel, C. (2017), 'PG-13 Movies contain more gun violence than R-Rated movies,' *Fast Company*, January 20.
Stolton, S. (2020a), 'Digital Services Act should avoid rules on 'harmful' content, Big Tech tells EU, Euractiv.com, 12 October,' https://www.euractiv.com/section/digital/news/digital-services-act-should-avoid-rules-on-harmful-content-big-tech-tells-eu/.
Stolton, S. (2020b), 'Content removal unlikely to be part of EU regulation on digital services, Jourova says,' *Euractiv.com*, September 23.
Taylor, R. (1999), 'Web auction sites face fraud inquiry,' *Financial Times*, January 27.
Teece, D. and H. Kahwaty (2021), 'Is the proposed digital markets act the cure for Europe's platform ills? Evidence from the European Commission's impact assessment,' https://lisboncouncil.net/wp-content/uploads/2021/04/TEECE-AND-KAHWATY-Dynamic-Digital-Markets1.pdf.
Travel Weekly. (2000), 'AMR spins off Sabre,' *Travel Weekly*, March 15.
Twitter. (2019), 'Rules Enforcement,' https://transparency.twitter.com/en/reports/rules-enforcement.html#2019-jul-dec.
U.K. Competition and Markets Authority. (2020), 'Online platforms and digital advertising: market study final report,' July 1. https://assets.publishing.service.gov.uk/media/5fa557668fa8f5788db46efc/Final_report_Digital_ALT_TEXT.pdf.
U.S. Congress. (1934), 'Communications Act of 1934,' https://transition.fcc.gov/Reports/1934new.pdf.
U.S. Congress. (1996), 'Telecommunications Act of 1996, U.S. Title 47, Section 230,' https://www.govinfo.gov/content/pkg/USCODE-2011-title47/pdf/USCODE-2011-title47-chap5-subchapII-partI-sec230.pdf.
U.S. Senate Committee on Commerce, Science, & Transportation. (2019), 'Mass violence, extremism, and digital responsibility, September 18,' https://www.commerce.senate.gov/2019/9/mass-violence-extremism-and-digital-responsibility.

U.S. Subcommittee on Antitrust, Commercial and Administrative Law of the Committee on the Judiciary. (2020), 'Investigation on competition in digital markets,' United States Government, https://judiciary.house.gov/uploadedfiles/competition_in_digital_markets.pdf.

Uberti, D. (2019), 'The German synagogue shooter's twitch video didn't go viral. Here's why,' *Vice News*, October 11.

von Mirbach, M. (1999), 'Demanding good wood,' in R. B. Gibson (ed), *Voluntary Initiatives: The New Politics of Corporate Greening*. Broadview Press: Waterloo, pp. 211–225.

Wells, G. (2020), 'Twitter says labels and warnings slowed spread of false election claims,' *Wall Street Journal*, November 12.

Wittern-Keller, L. (2008), *Freedom of the Screen*. The University Press of Kentucky: Lexington.

Wojcicki, S. (2017), 'Expanding our work against abuse of our platform,' YouTube Official blog post. https://youtube.googleblog.com/2017/12/expanding-our-work-against-abuse-of-our.html.

Woollacott, E. (2017), 'YouTube hires more moderators as content creators complain they're being unfairly targeted,' *Forbes*, December 5.

Zuboff, S. (2019), *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Profile Books: London.