A HOOVER INSTITUTION ESSAY

Fixing Social Media's Grand Bargain

JACK M. BALKIN

Aegis Series Paper No. 1814

To regulate social media in the twenty-first century, we should focus on its political economy: the nature of digital capitalism and how we pay for the digital public sphere we have. Our digital public sphere is premised on a grand bargain: free communications services in exchange for pervasive data collection and analysis. This allows companies to sell access to end users to the companies' advertisers and other businesses.

The political economy of digital capitalism creates perverse incentives for social media companies. It encourages companies to surveil, addict, and manipulate their end users and to strike deals with third parties who will further manipulate them.

Treating social media companies as public forums or public utilities is not the proper cure. It may actually make things worse. Even so, social media companies, whether they like it or not, have public obligations. They play important roles in organizing and curating public discussion and they have moral and professional responsibilities both to their end users and to the general public.

A reinvigorated competition law is one important way of dealing with the problems of social media, as I will describe later on. This essay, however, focuses on another approach: new fiduciary obligations that protect end-user privacy and counteract social media companies' bad incentives.

How Do We Pay for the Digital Public Sphere?

How does the political and economic system pay for the digital public sphere in our Second Gilded Age?[1] In large part, it pays for it through digital surveillance and through finding ever new ways to make money out of personal data.

Twenty-first-century social media like Facebook or YouTube differ from twentieth-century mass media like broadcast radio and television in two important respects. First, they are participatory, many-to-many media. Twentieth-century broadcast media are few-to-many: they publish and broadcast the content of a relatively small number of people to large audiences. In the twentieth century, most people would never get to use these facilities of mass communication to speak themselves. They were largely relegated to the role of audiences.

Twenty-first-century social media, by contrast, are many-to-many: they depend on mass participation as well as mass audiences. They make their money by encouraging enormous numbers of people to spend as much time as possible on their platforms and produce enormous amounts of content, even if that contribution is something as basic as commenting on, liking, or repeating somebody else's contribution. Facebook and Twitter would quickly collapse if people didn't constantly produce fresh content. Search engines, which are key parts of the digital infrastructure, also depend on people creating fresh links and fresh content that they can collect and organize.

Second, twenty-first-century social media like Facebook, YouTube, and Instagram rely on far more advanced and individualized targeted advertising than was available to twentieth-century broadcast media. Television and radio attempted to match advertisers with viewers, but there were limits to how finely grained they could target their audiences. (And newspapers, of course, relied on very broad audiences to sell classified advertisements.)

What makes targeted advertising possible is the collection, analysis, and collation of personal data from end users. Digital communication leaves collectible traces of interactions, choices, and activities. Hence digital companies can collect, analyze, and develop rich dossiers of data about end users. These include not only the information end users voluntarily share with others, but their contacts, friends, time spent on various pages, links visited, even keystrokes. The more that companies know about their end users, the more they know about other people who bear any similarity to them, even if the latter spend less time on the site or are not even clients. In the digital age, we are all constantly informing, not only on ourselves, but on our friends and relatives and, indeed, on everyone else in society.

This is not only true of social media, but of a wide range of digital services. The publisher of a paperback book in the 1960s could tell little about the reading habits of the people who purchased it, while Amazon can tell a great deal about the reading habits of the people who use their Kindle service, down to the length of time spent, the pages covered, the text highlighted and shared, and so on. As the Internet of things connects more and more devices and appliances to digital networks, surveillance spreads to ever more features of daily interaction. In general, the more interactive and the more social the service, the greater the opportunities for data collection, data analysis, and individualized treatment.

Data collection and analysis allow targeted advertising, which allows more efficient advertising campaigns, which allow greater revenues. But data collection and analysis offer another advantage: in theory, they give social media opportunities to structure and curate content for end users that they will find most engaging and interesting. That is important because advertising revenues depend on the amount of time and attention spent on the site. More engaging content means more time spent and more attention gained.

Social media companies have economic incentives to develop algorithms that will promote content that engages people. That is because companies' central goal is to gain attention share. This leads them to collect ever more data about their end users so that they can tailor content to individual end users to maximize their emotional engagement.[2]

This creates a problem. Often what engages people the most is material that produces strong emotional reactions—even if it is polarizing, false, or demagogic. Companies have economic incentives to expose people to this material. And unscrupulous actors, both domestic and foreign, have learned to take advantage of this feature of social media. As a result, the same business model that allows companies to maximize advertising revenues also makes them conduits and amplifiers for propaganda, conspiracy theories, and fake news.[3]

The Digital Grand Bargain and its Problems

Social media business models are a special case of the grand bargain that has made the digital public sphere possible in our Second Gilded Age. The bargain goes something like this: We will give you miraculous abilities. We will give you social media that allow you to connect with anyone, anywhere, anytime, in a fraction of a second. We will give you search engines that find anything you are looking for instantaneously. We will give you new forms of entertainment that are absorbing, engaging, outrageous, and amusing. We will give you ever more ways to measure yourself and express yourself to others.

We will give all of this to you, for free! And in return, you will let us surveil you. You will let us collect and analyze your habits, your locations, your links, your contacts with your friends, your mouse clicks, your keystrokes, anything we can measure. We will gladly take all of that and study it, and draw inferences from it, and monetize it, so that we can give you all these miraculous things. And we will use that data to perform experiments on you to figure out how to keep you even more focused on our sites and our products, so that you can produce even more data for us, which we can monetize.

This is the grand bargain of the Second Gilded Age. This is twenty-first-century data capitalism. And this is also the irony of the digital age: an era that promised unbounded opportunities for freedom of expression is also an era of increasing digital surveillance and control. The same technological advances allow both results. The infrastructure of digital free expression is also the infrastructure of digital surveillance.

What is objectionable about this grand bargain? The most obvious objection is that we must surrender individual privacy in order to speak. We must make ever more detailed portraits of our lives available to social media companies and their business partners. Beyond this, however, lies a deeper concern: the potential for abuse of power. In particular, the digital grand bargain creates an increased danger of manipulation—both by social media companies and by those who use social media companies—that is of a different degree and kind than

that which existed in the pre-digital era. By "manipulation" I mean techniques of persuasion and influence that (1) prey on another person's emotional vulnerabilities and lack of knowledge (2) to benefit oneself or one's allies and (3) reduce the welfare of the other person.[4] (Successful manipulation can also have ripple effects on third parties, such as family members and friends, or even fellow citizens.)

The problem with the current business models for social media companies such as Facebook, Twitter, and YouTube is that they give companies perverse incentives to manipulate end users—or to allow third parties to manipulate end users—if this might increase advertising revenues, profits, or both.

Manipulation is not a new problem. In the past, businesses have often appealed to people's emotions, desires, and weaknesses and have taken advantage of their relative ignorance. So have demagogues and political con artists. But the digital world of social media amplifies the opportunities for manipulation, both by social media companies and by those who use social media to reach end users.

The digital age exacerbates the twentieth-century problem of manipulation in several important respects. First, there is the issue of individual targeting. Twentieth-century influence campaigns were usually aimed at broad groups of individuals, with effects that were often hit-or-miss. With digital technologies it is now possible to tailor influence campaigns to individuals or to very small groups. Instead of appealing to the general emotional vulnerabilities of the public or the vulnerabilities of large demographic groups, digital companies can increasingly target the specific vulnerabilities and emotional hot buttons of individuals who may not be aware of precisely how they have been singled out.

Second, there are differences in scale, speed, and interactivity. Digital technologies allow individualized messages to be targeted to vast numbers of people simultaneously, something that was not possible with twentieth-century media. Moreover, end users' responses can be collected instantaneously, allowing companies to continually fine-tune their approaches, speeding up the Darwinian evolution of the most successful influence strategies. On top of this, digital companies now have the ability to perform interactive social science experiments on us to perfect their abilities to leverage and control our emotions. Facebook, for example, performed experiments to manipulate the emotional moods of 700,000 end users without their knowledge.[5] It has also experimented with ways of encouraging people to vote. But such techniques might also be used to discourage people from voting.[6] Moreover, these experiments can affect the behavior of not only end users but also those they come into contact with.[7]

Third, there is the problem of addiction. The more digital companies know about people's emotional vulnerabilities and predispositions, the more easily they can structure individual end-user experience to addict end users to the site.[8] Social media leverage the data they

collect about end users to offer periodic stimulation that keeps users connected and constantly checking and responding to social media. Media have always been designed to draw people's attention, but the digital experience can be especially immersive and pervasive, and thus a more powerful lure than a billboard or magazine advertisement. Here once again, the digital age far outstrips the powers of twentieth-century media.

One might object that, despite all this, the digital grand bargain remains freely chosen and welfare-enhancing. End users are free to use or not to use social media, and thus they are free to decide whether they will subject themselves to experimentation and emotional manipulation. If the free service is sufficiently valuable to them, they will accept the bargain. But this overlooks three important features of the emerging system of digital surveillance that make the assumption of a mutually beneficial arm's-length bargain highly implausible.

First, we cannot assume that transactions benefit both parties when there is extreme asymmetry of knowledge, in which one party's behaviors, beliefs, and activities are known to the other party while the other party is essentially a black box.

Second, individuals suffer from privacy myopia, a characteristic feature of digital interactions.[9] Individuals constantly generate a broad range of information about themselves through digital interactions, much of which (for example location, social connections, timing of responses, and rate of keystrokes) they may be only dimly aware. Individuals have no way of valuing or assessing the risks produced by the collection of particular kinds of information about them or how that information might be employed in the future. That is because the value of such information is cumulative and connective. Information that seems entirely irrelevant or innocuous can, in conjunction with other information, yield surprisingly powerful insights about individual values, behavior, desires, weaknesses, and predispositions. Because individuals cannot assess the value of what they are giving up, one cannot assume that their decisions enhance their welfare. In this environment, the idea of relying on informed consumer choice to discipline social media companies is a fantasy.

Third, as noted above, information gathered from end users has significant external effects on third parties who are not parties to the bargain. As digital companies know more about you, they also can learn more about other people who are similar to you or connected to you in some respect.[10] In the digital age, we do not simply inform on ourselves; we inform on other people as well. And when a social media company experiments with social moods or engineers an election, it affects not only its end users but many other people as well.

For all these reasons, it is fatuous to compare the digital grand bargain to a mutually beneficial arm's-length economic transaction. If we can pay for digital freedom of expression while reducing the dangers of digital manipulation, it is worth exploring alternatives.

Public Options

Proposals for reform of social media abound these days. One kind of proposal argues that we should counter the power of social media and search engines by treating them as state actors. Courts should apply standard First Amendment doctrine to them and treat them as public forums, which require complete content and viewpoint neutrality. If social media cannot choose what we see, they cannot manipulate us.

This solution fails to grapple with the central problems of the grand bargain. First, treating social media as public forums would only affect the ability of social media themselves to manipulate end users. It would do nothing to prevent third parties from using social media to manipulate end users, stoke hatred, fear, and prejudice, or spread fake news. And because social media would be required to serve as neutral public forums, they could do little to stop this. Second, even if social media do not curate feeds, they still collect end-user data. That end-user data, in turn, can be harvested and sold to third parties, who can use it on the site or elsewhere. (That is why, for example, requiring social media companies to offer a tiered service in which people pay not to receive commercial advertisements does not really deal with the underlying problem of surveillance, data collection, and manipulation.)

Perhaps equally important, the proposal is unworkable. Social media—and search engines—must make all sorts of editorial and curational judgments that the First Amendment forbids government entities to make.

For example, social media sites might want to require that end users use their real names or easily identifiable pseudonyms in order to limit trolling and abuse. They might decide to ban hate speech or dehumanizing speech, especially if they operate around the world. They might choose to ban graphic violence, nudity, or pornography. They might choose to ban advocacy of violence or illegal conduct, or the promotion of suicide. They might decide to ban certain types of harassment or incitement even if that harassment or incitement does not immediately lead to a breach of the peace.[11] They might ban certain forms of advertising. All of these regulations would be unconstitutional if a government imposed them in a public forum. More generally, we should accept that social media will have to make sometimes quite complicated decisions to discipline abusive trolls, maintain civility norms, demote the ranking of postings by conspiracy theorists and hate mongers, and, in cases of serial abuse, terminate accounts. Many of these policies would be unconstitutional if we applied the same standards to social media that the First Amendment applies to municipal streets and parks.

At a more basic level, it is impossible to manage a search engine or a social media site without curation, which involves a wide range of content-based judgments about what content to promote and what to demote.[12] It is also impractical and self-defeating to manage a social media site without moderation, which requires the imposition of a wide range of civility

rules that the First Amendment forbids governments from imposing in public discourse. Moreover, creating individualized social media feeds and search engine results inevitably requires content-based judgments. As described below, social media and search engines sometimes make bad decisions about these matters, but the solution is not to impose a set of doctrinal rules crafted for municipal streets and parks.

A second, related proposal argues that we should treat social media sites and search engines as public utilities because they perform what are clearly public functions. But public utility regulation—for example, of water and power utilities—generally focuses on two issues: access to essential services and fair pricing. Neither of these is particularly relevant. Social media and search engines want everyone to participate and they offer their services for free. If the goal of the public utility metaphor is to prevent content discrimination, it faces the same problems as treating digital media as state actors.

A third and quite different approach is public provisioning. Instead of treating existing private companies as arms of the state, governments could provide their own public options: government-run social media and search engines. For reasons stated above, these would not really work very well if they had to be organized as public forums and moderation was forbidden. There are potential solutions, however. The government could provide only a basic telecommunications system for social media messages and then allow various groups and businesses to create their own private moderation systems on top, from which individuals could choose. The government might also create an open system in which third parties could develop applications that allow people to design their own personalized feeds.

A government-provided search engine that is as efficient and effective as Google's is a somewhat harder lift and the cost of public provisioning for social media and search engines might be prohibitive. But public provisioning poses a far larger problem: state surveillance. Instead of Facebook and Google scooping up your personal data, the government would. The Fourth Amendment might not prohibit this under existing doctrines, because people willingly give the information to the public entity. Therefore any public provisioning system would have to be accompanied by very strict self-imposed restrictions on collection, analysis, and use. I am deeply skeptical that law enforcement and national security officials would willingly forgo access to all of this information.

Professional and Public-regarding Norms

We should not treat social media companies and search engines as state actors subject to the First Amendment. Yet we can still criticize them for arbitrariness and censorship. How is that possible if, as I have just explained, these companies must engage in content- and viewpoint-based judgments to do their jobs?

We can criticize social media companies in three ways, none of which requires us to treat them as state actors.

First, we can criticize them for being *opaque and non-transparent* and for denying basic norms of *fair process*. This happens when social media do not state their criteria for governance clearly in advance and do not offer reasoned explanations for their decisions.

Second, we can criticize them for being *arbitrary*—for not living up to their own community guidelines and terms of service. They should apply their own rules without fear or favor to the rich and to the poor, the high and low alike. Twitter and Facebook, to name two examples, have often been lax with violations of their terms of service by famous or well-known people and strict with violations by people who are not famous or well known.[13] This allows the more powerful and famous to abuse the less powerful with impunity and it creates blind spots in enforcement.

Third, and perhaps more important, we can criticize social media companies for failing to live up to norms of *professionalism* and *expertise*—that is, for failing to live up to the norms of the kind of entity they purport to be.

Here is an analogy. People criticize major newspapers and media outlets all the time. They criticize them for biased coverage, they criticize them for giving a platform to people who make stupid or evil arguments, and they criticize them for failing to educate the public about the issues of the day.

In most cases, people understand that these criticisms aren't supposed to lead to government regulation of newspapers and mass media. People understand that these companies have a First Amendment right to exercise editorial discretion as they see fit, even if they exercise it badly. Nevertheless, they hold these companies to a higher standard than ordinary individuals expressing their opinions. The public rightly assumes that media companies should live up to certain professional standards that are both public-regarding and connected to democratic life. These include, among other things, providing the public with important information necessary to self-government, striving to cover the news accurately and fairly, engaging in professional fact-checking, adhering to professional standards of journalistic ethics, and so on.

Many media organizations fail to live up to these standards, often spectacularly so. And some media organizations have essentially given up on professionalism, fairness, and accuracy. But people generally understand that this is a valid reason to criticize them, not to exculpate them. Media companies hold themselves out as adhering to professional and public-regarding norms. Therefore people in a democracy feel that they have a right to criticize them when, in their estimation, media fail to live up to those norms. Perhaps equally important, because the norms are public-regarding, citizens in a democracy feel that they have a right to debate

what those professional norms should be, whether or not the media companies assume them or live up to them.

Social media companies and search engine companies are not newspapers. Even so, they are more than just run-of-the-mill companies. They do more than just serve ads or sell widgets. They perform a public service—three connected services, in fact. First, they *facilitate* public participation in art, politics, and culture. Second, they *organize* public conversation so that people can easily find and communicate with each other. Third, they *curate* public opinion through individualized results and feeds and through enforcing terms-of-service obligations and community guidelines.

These digital companies are the twenty-first-century successors of twentieth-century mass media companies, even though their functions are somewhat different. The public, not surprisingly, has come to view them as having a public-oriented mission.

In fact, these companies encourage this understanding through the ways they talk about themselves. The Twitter Rules, for example, begin with the statement, "We believe that everyone should have the power to create and share ideas and information instantly, without barriers. In order to protect the experience and safety of people who use Twitter, there are some limitations on the type of content and behavior that we allow."[14] This is a statement of public-regarding, professional norms for facilitating public participation, organizing public discussion, and curating public opinion. Facebook and YouTube have made similar statements of purpose and justifications for their community guidelines, although their policies differ in some respects.[15]

Whether they imagined it or not at the outset, these companies have taken on a public function. People may therefore criticize them—and should criticize them—if they feel that these companies are acting contrary to appropriate professional norms.

Moreover, because these companies have taken on these three tasks—facilitating public participation, organizing public conversation, and curating public opinion—they may also impose basic civility norms against abuse, threats, and harassment. They may also ban hate speech or speech that denigrates people if they think that this kind of speech will undermine the public-regarding purposes of the site. Social media companies may do this even if the First Amendment would prevent the federal government from imposing the same civility norms on a government-operated social media site.

But if social media companies decide to govern their sites through imposing civility norms and regulating harassment and abuse, they should abide by the two other basic norms stated above. First, they should be transparent about what they are doing and why they are doing it. Second, they should not be arbitrary in their governance.

Social media companies have been only fitfully successful at meeting these obligations. Understood charitably, we might say that they are at the very beginning of a long process of learning how to be responsible professionals. They have been wildly successful as technology companies, but professionalism is more than technological expertise. Professional judgments may require the application of norms that do not scale well. Sometimes applying these norms will require judgment and individualized assessment as well as algorithmic sorting and bright-line rules. Doing this costs more in human resources and attention than purely technological solutions. To the extent that this is the case, social media companies should absorb the extra costs of being professionals and living up to professional norms. Although their efforts have been halting and often inadequate, social media companies are slowly beginning that arduous process. In the meantime, civil society can play an important role by continuing to criticize social media companies and by encouraging them to live up to their public responsibilities.

Reforming Social Media

I have already said that we should not use the law to force these companies to behave as public-regarding professionals any more than we can force major newspapers to adhere to proper journalistic standards. Does this mean that law has no role to play? No. The law may encourage these public-regarding norms in certain limited ways consistent with the First Amendment.

Instead of directly aiming at the editorial policies of social media companies, reform proposals should focus instead on the grand bargain that has turned the infrastructure of digital free expression into the infrastructure of digital surveillance and control. Social media companies will continue to cause a host of social problems as long as their business models cause them not to care about these problems.

There are two central ways to change their behavior. The first is to reshape the organization of social media companies. This is the task of antitrust and pro-competition law, which have grown moribund in the Second Gilded Age and need a serious rethinking.

Social media companies' perverse incentives derive from their business models—selling end users' information to advertisers and manipulating and addicting end users so that they spend more time on social media and are thus more accessible to advertisers. Because a small number of social media dominate end users' attention, they also have a stranglehold over digital advertising. People who wish to advertise online must operate primarily through Facebook's and Google's advertising networks. This reduces revenues for many news and media sites that are crucial to the health and vibrancy of the digital public sphere.

Increased enforcement of existing antitrust laws and a series of new pro-competition policies might have two salutary effects. First, these reforms might restructure how digital advertising

operates, ameliorating the current bottleneck and freeing up revenues for a wider range of media companies. Second, reform of competition policy and stricter antitrust enforcement might break up the largest companies into smaller companies that can compete with each other or create a space for new competitors to emerge. (Facebook and Google have often bought up potential competitors before they could grow large enough to threaten them.)

More social media companies mean more platforms for innovation and more different software features and affordances. More companies might also make it more difficult for foreign hackers to disrupt the digital public sphere. All other things being equal, it may be harder to hack twelve Facebooks than only one.[16] Finally, more different kinds of companies might also provide more models for social spaces and communities and a wider variety of speech policies.

This last point is especially important. I have just argued that social media companies must be allowed to enforce civility norms and regulate or even ban a wide range of speech that state actors may not touch. But modern democracies increasingly rely on social media to perform the public functions of organizing public opinion and facilitating public discussion. Therefore it is very important to ensure that there are many social media applications and businesses in order to prevent a small number of powerful for-profit companies from dominating how public opinion is organized and governed.

Moreover, social media companies often enforce their terms of service imperfectly and arbitrarily and they may make many questionable judgments. Some, like Facebook, attempt to impose the same standards around the world.[17] Finally, civil society organizations, mass media, politicians, and governments have and will put increasing pressure on social media to ban speech that they do not like and expel speakers who offend them. All of them, in various ways, will try to coax social media into serving their political or ideological agendas. These are all reasons for using pro-competition laws to ensure a healthy number of competing firms organizing public discourse. Precisely because people will demand that huge multinational corporations ban speech they do not like, it is important to have many Facebooks, not just one. If we expect social media sites to enforce civility norms, we also need multiple social media sites serving different values and different publics.

Information Fiduciaries

A second approach to reform is to make social media companies internalize the costs they impose on society through surveillance, addiction, and manipulation by giving them new social responsibilities. The short-term goal is to counteract the most egregious examples of bad behavior. The long-term goal is to create legal incentives for social media companies to develop professional cultures and public-oriented norms for organizing and curating public discussion. To do this, I propose reaching back to some very old ideas in the law that governs the professions: namely, the idea of fiduciary obligation.

We should treat social media companies—and many other digital media companies as well—as *information fiduciaries* toward their clients and end users.[18] As information fiduciaries, digital companies should have duties of care, confidentiality, and loyalty toward the people whose data they collect, store, and use. This reform is a natural outgrowth of the grand bargain that has enabled free expression in the digital age.

Because of digital companies' increasing capacities for surveillance and control, they must take on new legal responsibilities. Put simply, digital companies know a lot about us, and they can use that knowledge in many ways—but we don't know a lot about them. Moreover, people increasingly depend on a wide range of digital services that observe them and collect data about them. That makes people increasingly vulnerable to these companies. Because the companies' operations are not transparent, people have to trust that these services will not betray them or manipulate them for their own ends. Digital companies that create and maintain this dependence and vulnerability should be considered information fiduciaries toward their end users.

There is plenty of precedent for this idea. For centuries, the law has recognized that certain people hold power over others who are vulnerable to them, dependent on them, and have to trust them. It created the idea of fiduciary obligations for just these situations.[19] For example, the law has long maintained that the clients or patients of doctors and lawyers are in special relationships of dependence and vulnerability. We need to trust these professionals with sensitive personal information about ourselves, but the people we trust could use this same information to harm us and enrich themselves in many different ways. Therefore the law treats professionals like doctors, lawyers, accountants, and estate managers as fiduciaries. Fiduciary relationships require good faith and loyalty toward people whom the relationships place in special positions of vulnerability. Accordingly, fiduciaries have special duties of care, confidentiality, and loyalty toward their clients and beneficiaries.

Because social media companies collect so much data about their end users, use that data to predict and control what end users will do, and match them with third parties who may take advantage of end users, they are among the most important examples of the new information fiduciaries of the digital age. We should apply these traditional obligations to the changed conditions of a new technological era.

Facebook is not your doctor or lawyer. YouTube is not your accountant or estate manager. We should be careful to tailor the fiduciary obligations to the nature of the business and to the reasonable expectations of consumers. That means that social media companies' fiduciary duties will be more limited.

Social media companies and search engines provide free services in exchange for the right to collect and analyze personal data and serve targeted ads. This by itself does not violate fiduciary obligations. Nevertheless, it creates a perpetual conflict of interest between end

users and social media companies. Companies will always be tempted to use the data they collect in ways that increase their profits to their end users' disadvantage. Unless we are to ban targeted advertising altogether (which I would oppose and which raises serious First Amendment problems) the goal should be to ameliorate or forestall conflicts of interest and impose duties of good faith and non-manipulation. That means that the law should limit how social media companies can make money off their end users, just as the law limits how other fiduciaries can make money off their clients and beneficiaries.

As information fiduciaries, social media companies have three major duties: duties of care, duties of confidentiality, and duties of loyalty. The duties of care and confidentiality require fiduciaries to secure customer data and not disclose it to anyone who does not agree to assume similar fiduciary obligations. In other words, fiduciary obligations must run with the data. The duty of loyalty means that fiduciaries must not seek to advantage themselves at their end users' expense and they must work to avoid creating conflicts of interest that will tempt them to do so. At base, the obligations of loyalty mean that digital fiduciaries may not act like con artists. They may not induce trust on the part of their user base and then turn around and betray that trust in order to benefit themselves.

To see what these obligations would mean in practice, we can use the Cambridge Analytica scandal that propelled the issue of social media regulation to public attention in the spring of 2018.

Although the facts are complicated, they essentially involved Facebook's decision to allow third parties to access its end users' data.[20] Facebook allowed researchers to do this for free and took a cut of the profits for business entities. This allowed it to leverage its central resource—consumer data—to increase profits.

Aleksandr Kogan, a data scientist, used a personality quiz to gain access to Facebook's end-user data. He thereby obtained not only the data of the 300,000 people who logged in using their Facebook credentials, but also all of their Facebook friends, an estimated 87 million people.[21] In fact, Kogan was actually working for Cambridge Analytica, a for-profit political consulting company. Cambridge Analytica used the end-user data to produce psychological profiles that, in turn, it would use to target political advertisements to unsuspecting Facebook users. In fact, these practices were only the tip of a far larger iceberg. Facebook made a series of unwise decisions to allow a range of business partners access to its end users' social graphs and thus make them vulnerable to various kinds of manipulation.[22]

As an information fiduciary, Facebook violated all three of its duties of care, confidentiality, and loyalty. It did not take sufficient care to vet its academic and business partners. It did not ensure that it only gave access to data to entities that would promise to maintain the same duties of care, confidentiality, and loyalty as Facebook. It did not take sufficient steps to audit and oversee the operations of these third parties to ensure that they did not violate

the interests of its end users. It allowed third parties to manipulate its end users for profit. And when it discovered what had happened, many years later, it did not take sufficient steps to claw back its end users' data and protect them from further breaches of confidentiality and misuse.

Fiduciary obligations matter most in situations in which social media companies have powerful market incentives not to protect their end users: for example, when social media companies give access to data to third-party companies without adequate safeguards to prevent these third parties from manipulating end users. Fiduciary obligations also matter when social media companies perform social science experiments on their end-user base. Social media companies are not part of universities and therefore are not bound by human-subjects research obligations. As information fiduciaries, however, they would have legal duties not to create an unreasonable risk of harm to their end users or to the public for their own advantage. They would have duties, just as university scientists do, to minimize harm and prevent overreaching and manipulation by their employees and contractors.

Finally, if social media companies are information fiduciaries, they should also have a duty not to use end-user data to addict end users and psychologically manipulate them. Social media companies engage in manipulation when end users must provide information in order to use the service and when companies use this information to induce end-user decision making that benefits the company at the expense of the end user and causes harm to the end user. Because this creates a conflict of interest between the company and its end users, it violates the duty of loyalty.

It may be useful to compare the fiduciary approach with the privacy obligations of the European Union's General Data Protection Regulation (GDPR). There is considerable overlap between the two approaches. But the most important difference is that the GDPR relies heavily on securing privacy by obtaining end-user consent to individual transactions. In many respects, it is still based on a contractual model of privacy protection. Contractual models will prove insufficient if end users are unable to assess the cumulative risk of granting permission and therefore must depend on the good will of data processors. The fiduciary approach to obligation does not turn on consent to particular transactions, nor is it bounded by the precise terms of a company's written privacy policy or terms of service, which are easy for companies to modify. Rather, the fiduciary approach holds digital fiduciaries to obligations of good faith and non-manipulation regardless of what their privacy policies say.

The fiduciary approach is also consistent with the First Amendment. That is because it aims at regulating the relationships of vulnerability and trust between information fiduciaries and those who must trust them.[23]

The First Amendment treats information gained in the course of a fiduciary relationship differently from other kinds of information. Tell a secret to a person in the street and he

or she can publish it tomorrow and even use it against your interests. But when you reveal information to a fiduciary—a doctor, nurse, or lawyer—he or she has to keep it confidential and cannot use it against you. Information gained in the course of a fiduciary relationship—and that includes the information that social media companies collect about us—is not part of the public discourse that receives standard First Amendment protection. Instead, the First Amendment allows governments to regulate fiduciaries' collection, collation, use, and distribution of personal information in order to prevent overreaching and to preserve trust and confidentiality.[24] The same principle should apply to the new information fiduciaries of the digital age.

There may be close cases in which we cannot be sure whether a company really is acting as an information fiduciary. To deal with these situations, Jonathan Zittrain and I have proposed that Congress offer digital companies a different grand bargain to protect end users' privacy.[25] It would create a safe harbor provision for companies that agree to assume fiduciary obligations. The federal government would preempt state regulation if digital media companies accept the obligations of information fiduciaries toward their end users. Offering this exchange does not violate the First Amendment.

For the most part, the fiduciary approach leaves social media companies free to decide how they want to curate and organize public discussion, focusing instead on protecting privacy and preventing incentives for betrayal and manipulation. It affects companies' curation and organization of public discourse only to the extent that companies violate their duties of care, confidentiality, and loyalty.

The fiduciary approach has many advantages. It is not tied to any particular technology. It can adapt to technological change. It can be implemented at the state or the federal level, and by judges, legislatures, or administrative agencies.

The fiduciary approach also meshes well with other forms of consumer protection, and it does not exclude other reforms, like GDPR-style privacy regulation. In particular, it does not get in the way of new pro-competition rules or increased antitrust enforcement as described above. That is because it does not turn on the size of an organization (although Congress might choose to regulate only the larger sites in order to encourage innovation and avoid barriers to entry). It also does not turn on the presence or absence of monopoly power. It applies whether we have twelve Facebooks or only one. Indeed, even if we had a wide range of social media companies, all harvesting, analyzing, and using end-user data, there would still be a need for fiduciary obligations to prevent overreaching.

The fiduciary approach pays attention to deeper causes. It directs its attention to the political economy of digital media. It focuses on repairing the grand bargain that pays for the digital public sphere in the Second Gilded Age.

NOTES

1 During the First Gilded Age, which ran from the end of Reconstruction to the beginning of the twentieth century, technological innovation created huge fortunes in the hands of a small number of entrepreneurs and produced increasing inequalities of wealth and deep political corruption. Waves of immigration and increasing racial tensions led to the emergence of populist demagogues. American government was increasingly for sale, and many people despaired for the future of American democracy. The corruption and inequality of the First Gilded Age led to the reforms of the Progressive Era and, eventually, the New Deal. For a general history of the period, see H. W. Brands, *American Colossus: The Triumph of Capitalism, 1865–1900* (New York: Anchor, 2010).

2 See Zeynep Tufecki, "Facebook's Surveillance Machine," *New York Times*, March 19, 2018, accessed September 27, 2018, https://www.nytimes.com/2018/03/19/opinion/facebook-cambridge-analytica.html. ("Facebook makes money, in other words, by profiling us and then selling our attention to advertisers, political actors and others. These are Facebook's true customers, whom it works hard to please.") These business models and the incentives they create are examples of what Shoshana Zuboff calls "surveillance capitalism." Shoshana Zuboff, "Big Other: Surveillance Capitalism and the Prospects of an Information Civilization," *Journal of Information Technology* 30 (April 2015): 75 (defining "surveillance capitalism" as a "new logic of accumulation" and a "new form of information capitalism [that] aims to predict and modify human behavior as a means to produce revenue and market control").

3 See, e.g., Paul Lewis, "'Fiction Is Outperforming Reality': How YouTube's Algorithm Distorts Truth," *Guardian*, February 2, 2018, accessed September 27, 2018, https://www.theguardian.com/technology/2018/feb/02/how-youtubes-algorithm-distorts-truth (explaining how YouTube's algorithm to engage viewers promotes conspiracy theories).

4 This definition of manipulation focuses on leveraging another's lack of knowledge and emotional vulnerability to benefit oneself at the expense of another's welfare. This is not the only way to define the concept. See, e.g., Cass R. Sunstein, *The Ethics of Influence: Government in the Age of Behavioral Science* (New York: Cambridge University Press, 2016), 82 (a technique of influence is "manipulative to the extent that it does not sufficiently engage or appeal to [a person's] capacity for reflection and deliberation"). That definition, however, raises the problem of how to distinguish manipulation from a wide range of ordinary techniques of marketing. A third approach would focus on real or objective interests: manipulation is persuasion that leverages lack of knowledge and emotional vulnerability to cause people to act against their real interests, however those are defined. This approach raises the question of how we know what people's real or objective interests are.

5 "Facebook Admits Failings over Emotion Manipulation Study," BBC News, October 3, 2014, accessed September 27, 2018, https://www.bbc.com/news/technology-29475019 ("the company was widely criticised for manipulating material from people's personal lives in order to play with user emotions or make them sad").

6 Jonathan Zittrain, "Engineering an Election," *Harvard Law Review Forum* 127 (June 20, 2014): 335–36 (noting that experiment caused an additional 340,000 votes to be cast).

7 Ibid., 336 (describing the "ripple effects" of experiments).

8 See Mike Allen, "Sean Parker Unloads on Facebook: 'God Only Knows What It's Doing to Our Children's Brains,'" *Axios*, November 9, 2017, accessed September 27, 2018, https://www.axios.com/sean-parker-unloads-on-facebook-god-only-knows-what-its-doing-to-our-childrens-brains-1513306792-f855e7b4-4e99-4d60-8d51-2775559c2671.html (quoting statement by former president of Facebook that social media applications are designed to "exploit a vulnerability in human psychology" using psychological methods to "consume as much of your time and conscious attention as possible" and keep users locked into the site); Paul Lewis, "'Our Minds Can Be Hijacked': The Tech Insiders who Fear a Smartphone Dystopia," *Guardian*, October 6, 2017, accessed September 27, 2018, https://www.theguardian.com/technology/2017/oct/05/smartphone-addiction-silicon-valley-dystopia (interviewing former employees at Google and Facebook who report that technologies are designed to addict users and monopolize their attention).

9 See, e.g., Ryan Calo, "The Boundaries of Privacy Harm," *Indiana Law Journal* 86, no. 3 (October 4, 2011): 1131, 1149 ("Many consumers have little idea how much of their information they are giving up or how it will be used"); A. Michael Froomkin, "The Death of Privacy?" *Stanford Law Review* 52 (2000): 1461, 1502 ("Consumers suffer from privacy myopia: they will sell their data too often and too cheaply"); Daniel J. Solove, "Privacy and Power: Computer Databases and Metaphors for Information Privacy," *Stanford Law Review* 53 (2001): 1393, 1452 ("It is difficult for the individual to adequately value specific pieces of personal information").

10 See Tufekci, "Facebook's Surveillance Machine," explaining that Facebook collects "shadow profiles" on nonusers: "even if you are not on Facebook, the company may well have compiled a profile of you, inferred from data provided by your friends or from other data. This is an involuntary dossier from which you cannot opt out in the United States." Social media users may unwittingly imperil each other's privacy. The Cambridge Analytica scandal revealed that when Facebook users logged in to a third-party app using their Facebook credentials, they shared the social graphs of all of their Facebook friends without the latter's consent. See Alexandra Samuel, "The Shady Data-Gathering Tactics Used by Cambridge Analytica Were an Open Secret to Online Marketers. I Know, Because I Was One," *The Verge*, March 25, 2018, accessed September 27, 2018, https://www.theverge.com/2018/3/25/17161726/facebook-cambridge-analytica-data-online-marketers. ("The tactic of collecting friend data, which has been featured prominently in the Cambridge Analytica coverage, was a well-known way of turning a handful of app users into a goldmine.")

11 For examples of what social media sites regulate, see Facebook, Community Standards, https://www.facebook.com/communitystandards; and Twitter, The Twitter Rules, https://help.twitter.com/en/rules-and-policies/twitter-rules (both accessed September 27, 2018).

12 Tarleton Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media* (New Haven, CT: Yale University Press, 2018).

13 See Kate Klonick, "The New Governors: The People, Rules, and Processes Governing Online Speech," *Harvard Law Review* 131 (April 10, 2018): 1598, 1654–55 (2018) (noting that social media companies may disproportionately favor people with power over other end users).

14 Twitter, the Twitter Rules.

15 Facebook, Community Standards, "We recognize how important it is for Facebook to be a place where people feel empowered to communicate, and we take our role in keeping abuse off our service seriously. That's why we have developed a set of Community Standards that outline what is and is not allowed on Facebook. . . . The goal of our Community Standards is to encourage expression and create a safe environment," YouTube, Policies and Safety, accessed September 27, 2018, https://www.youtube.com/yt/about/policies/#community-guidelines, "When you use YouTube, you join a community of people from all over the world. . . . Following the guidelines below helps to keep YouTube fun and enjoyable for everyone."

16 Sally Hubbard, "Fake News is a Real Antitrust Problem," *CPI Antitrust Chronicle*, December 2017: 5, accessed September 27, 2018, https://www.competitionpolicyinternational.com/wp-content/uploads/2017/12/CPI-Hubbard.pdf.

17 Klonick, "New Governors," 1642, describing Facebook's goal of applying its norms worldwide and the resulting compromises; Julia Angwin and Hannes Grassegger, "Facebook's Secret Censorship Rules Protect White Men from Hate Speech but Not Black Children," *ProPublica*, June 28, 2017, accessed September 27, 2018, https://www.propublica.org/article/facebook-hate-speech-censorship-internal-documents-algorithms (describing Facebook's attempts to enforce its hate speech rules worldwide and the arbitrariness of its categories).

18 See Jack M. Balkin, "Information Fiduciaries and the First Amendment," *UC Davis Law Review* 49, no. 4 (April 2016): 1183; Jack M. Balkin, "The Three Laws of Robotics in the Age of Big Data," *Ohio State Law Journal* 78 (2017): 1217.

19 See generally Tamar Frankel, *Fiduciary Law* (New York: Oxford University Press, 2011).

20 See Carole Cadwalladr and Emma Graham-Harrison, "How Cambridge Analytica Turned Facebook 'Likes' into a Lucrative Political Tool," *Guardian*, March 17, 2018, accessed September 27, 2018, https://www.theguardian.com/technology/2018/mar/17/facebook-cambridge-analytica-kogan-data-algorithm; Carole Cadwalladr and Emma Graham-Harrison, "Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach," *Guardian*, March. 17, 2018, accessed September 27, 2018, https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election; Paul Lewis, "'Utterly Horrifying': Ex-Facebook Insider Says Covert Data Harvesting Was Routine," *Guardian*, March 20, 2018, https://www.theguardian.com/news/2018/mar/20/facebook-data-cambridge-analytica-sandy-parakilas.

21 See Michael Riley, Sarah Frier, and Stephanie Baker, "Understanding the Facebook-Cambridge Analytica Story: QuickTake," *Washington Post*, April 9, 2018, accessed September 27, 2018, https://www.washingtonpost.com/business/understanding-the-facebook-cambridge-analytica-story-quicktake/2018/04/09/0f18d91c-3c1c-11e8-955b-7d2e19b79966_story.html (estimating that 300,000 people participated and that 87 million users had their data harvested).

22 Lewis, "Covert Data Harvesting Was Routine" (quoting a former Facebook employee who explained that under the company's policies, "a majority of Facebook users" could have had their data harvested by app developers without their knowledge).

23 On the First Amendment issues, see Balkin, "Information Fiduciaries and the First Amendment," 6.

24 Ibid.

25 Jack M. Balkin and Jonathan Zittrain, "A Grand Bargain to Make Tech Companies Trustworthy," *Atlantic*, October 3, 2016, accessed September 27, 2018, https://www.theatlantic.com/technology/archive/2016/10/information-fiduciary/502346.

The preferred citation for this publication is: Jack Balkin, “Fixing Social Media’s Grand Bargin,” Hoover Working Group on National Security, Technology, and Law, Aegis Series Paper No. 1814 (October 16, 2018), available at https://www.lawfareblog.com/advanced-persistent-manipulators-and-social-media-nationalism-national-security-world-audiences.

About the Author

JACK M. BALKIN

Jack M. Balkin is Knight Professor of Constitutional Law and the First Amendment at Yale Law School and the founder and director of Yale's Information Society Project. He is a member of the American Academy of Arts and Sciences and founded and edits the group blog Balkinization (http://balkin.blogspot.com/). His most recent book is *Living Originalism* (Cambridge, MA: Belknap Press, 2011).

Working Group on National Security, Technology, and Law

The Working Group on National Security, Technology, and Law brings together national and international specialists with broad interdisciplinary expertise to analyze how technology affects national security and national security law and how governments can use that technology to defend themselves, consistent with constitutional values and the rule of law.

The group focuses on a broad range of interests, from surveillance to counterterrorism to the dramatic impact that rapid technological change—digitalization, computerization, miniaturization, and automaticity—are having on national security and national security law. Topics include cybersecurity, the rise of drones and autonomous weapons systems, and the need for—and dangers of—state surveillance. The group's output will also be published on the Lawfare blog, which covers the merits of the underlying legal and policy debates of actions taken or contemplated to protect the nation and the nation's laws and legal institutions.

Jack Goldsmith and Benjamin Wittes are the cochairs of the National Security, Technology, and Law Working Group.

For more information about this Hoover Institution Working Group, visit us online at http://www.hoover.org/research-teams/national-security-technology-law-working-group.

Hoover Institution, Stanford University
434 Galvez Mall
Stanford, CA 94305-6003
650-723-1754

Hoover Institution in Washington
The Johnson Center
1399 New York Avenue NW, Suite 500
Washington, DC 20005
202-760-3200

HOOVER INSTITUTION