Adapting to the New Public Forum: Free Expression in the Age of Big Tech

Brendin Flinn

Dan O'Hanlon Essay Competition, 2021

Abstract

The advent of the internet—and the increased ease and breadth of communication that it has afforded—has drastically altered the communication landscape. Social media platforms have taken center stage in conversations about how internet communication has altered public discourse due to their ability to connect geographically distant users with just a few clicks. Indeed, social media platforms can rightly be viewed as the modern equivalent of traditional public forums (parks, public buildings, sidewalks, etc.). Despite all of the benefits that can be attributed to these new public forums, various criticisms have been levied against unique aspects of social media platforms and how they choose to operate. Some of the most commonly expressed criticisms pertain to the unique influences these platforms have (or could have) on free expression and public discourse. Such criticisms have been used to support the idea that these platforms require legislative reforms, particularly (for the United States) reforms to section 230 of the Communications Decency Act of 1996. This essay explores free expression—especially how it has changed over time with changing attitudes and technology—and argues for reforming both section 230 and the practices of social media platforms with the aim of ensuring that free expression remains robust in the face of the unique challenges confronting it in the internet age. Specifically, this essay argues that treating social media platforms as common carriers could shield free expression from undue restrictions.

Table of Contents

Section I: Traditional Understandings and Evolution of Free Expression

A: Basic Principles Undergirding Free Expression

Understanding the current state of affairs necessitates starting from the basic principles with which we make judgements about the freedom to express oneself. Such a starting point—in the context of modern liberal democracies at least—can be found in the writings of John Stuart Mill, particularly in *On Liberty*. Mill argued for an incredible degree of freedom with respect to speech and free expression: "If all mankind minus one were of one opinion, and only one person were of the contrary opinion, mankind would be no more justified in silencing that one person than he, if he had the power, would be justified in silencing mankind" (Mill, 1859). Importantly, Mill didn't only argue for such a profound ethical restriction on limiting free expression for ideas which society generally finds to be benign: "If the arguments of the present chapter are of any validity, there ought to exist the fullest liberty of professing and discussing, as a matter of ethical conviction, any doctrine, however immoral it may be considered" (Mill, 1859). Over the course of time, this viewpoint has largely remained intact. Noam Chomsky, for instance, pointed out that protections for free expression apply most directly to those ideas which society deems to be undesirable, fallacious, or immoral because there is little need to protect speech that most people would not wish to see restricted (Chomsky, 1980). This idea was also present in the minds of those who enshrined the freedom of expression in the First Amendment to the United States Constitution (James Madison in particular). The ability to speak freely—especially to speak ideas which were contrary to public sentiments—was viewed not merely as a right afforded to the people, but was rather viewed as a necessity to maintain the sovereignty of the people in a free society (Cost, 2017).

The general justifications for maintaining freedom of expression in democracies can be broken down into two categories: procedural and substantive (Nunziato, 2003). Procedural justifications argue that freedom of expression is necessary to the very process of democracy; in other words, democracy cannot function properly unless citizens are afforded the right to speak freely (Nunziato, 2003). While freedom of expression is not the only right that procedural democratic theorists argue to be integral to democratic processes, it takes center stage because democratic processes rest on thoughtful discussion between members of a democratic society. For representative forms of democracy, communication between representatives and their constituents is also a necessary precondition to ensuring that self-governance is exercised—if representatives do not know the wishes of the people, then the wishes of the people cannot shape governance. Substantive justifications take a different approach. Rather than asserting that free expression must be protected because self-governance is desirable, substantive justifications rest on the idea that democracies can only form when citizens are viewed as equal, autonomous, and rational agents, thus necessitating that the government allow them to exercise their will through speech (Nunziato, 2003). This derivation of the freedom to express oneself in a democratic society has been forwarded by the likes of Immanuel Kant and John Rawls.

Despite the belief in the necessity of free expression, it has not been viewed to be absolute. Indeed, Mill's proposition that even the most fringe beliefs ought to be allowed to be judged by the public conscience was contrasted by his assertion that the freedom of expression can be reasonably limited when an expression is likely to induce harm to another individual or individuals (Mill, 1859). This libertarian concept, coined as the 'harm principle', is one of the most widely accepted tools employed by liberal thinkers to determine ethical restrictions on the freedom of expression (van Mill, 2021). In fact, such a restriction is necessary to maintain

consistency with the principles that are used to derive the freedom of expression in the first place. A value often used to argue for the inherent freedom of expression is autonomy, but if it is the case that expression can be used in ways to impose on the autonomy of others, then restrictions on that expression are demanded (van Mill, 2021). More broadly, we would do well to keep in mind that the freedom of expression is situated in the context of other, potentially competing, social values. We value speech as highly as we do because of its benefits to certain social values. If free expression accomplished little that we value, then we would care little about restrictions placed upon it. Considering this, our endeavors to understand reasonable limits on free expression (using the harm principle, for instance) entail determinations of the implications of certain expressions for other social values. All free societies place certain restrictions on free expression (to varying degrees) as a result of the ability for certain expressions to bring harm upon competing values, security being the most common (van Mill, 2021). Thus, our objective should never be to reject restrictions on free expression *prima facie*. Instead, we ought to aim to ensure that free expression is not *unreasonably* restricted in cases where it comes into conflict with competing values. This is not a simple task, but it is a necessary one.

B: Evolution of First Amendment Free Expression in Supreme Court Case Law

While understanding the basic principles foundational to free expression is necessary to our questioning of modern practices, it is not alone sufficient to answer such questions; free expression is situated within the legal realm, making analysis of the evolution of our legal understanding of free expression necessary to both understanding the status quo and progressing from it. Within the United States, our conceptualization of free expression was first outlined in the First Amendment to the Constitution but has been expanded repeatedly by interpretations of

that amendment (as well as its relation to other amendments) by the Supreme Court of the United States.

Though it may seem odd to many today, the Supreme Court's initial rulings regarding the first Ten Amendments—of note, *Barron v. Baltimore* (1833)—concluded that the amendments applied only to the federal government, not state or local governments. This thought persisted, and was utilized in a variety of cases, until the Court's ruling in *Gitlow v. New York* (1925) utilized the incorporation principle to justify the application of certain prohibitions in the First Ten Amendments (in this case, free speech) to the states. Prior to *Gitlow*, the Court's interpretations regarding free expression under the First Amendment were limited to federal cases. Interestingly, the first significant cases involving free expression taken up by the Court did not appear until the 19th century, just before (and including) *Gitlow*. Though likely surprising to many today, these first cases largely sided in favor of restricting certain forms of expression. The White Court's decisions in *Schenck v. United States* (1919), *Abrams v. United States* (1919), and *Debs v. United States* (1919) asserted that restrictions on speech that undermined national security (these cases involved speech determined to undermine the war effort during World War I) were allowed under the First Amendment. Two important notes from these cases are the establishment of the "clear and present danger" test (articulated in Justice Holmes' opinion for the unanimous decision in *Schenck* and used again in both *Abrams* and *Debs*) and the broad determination that the government should be afforded greater deference in wartime. Perhaps of greater note, however, is the dissenting opinion of Justice Holmes in *Abrams*. Not only did Justice Holmes argue that the dissemination of leaflets condemning federal wartime actions was a constitutionally protected expression of political dissent, but he also invoked Mill's concept of the competition of competing claims in a free market (Cohen, 2013). Indeed, Holmes' powerful

dissent (along with his future endeavors on the Court) would go on to help restrict federal power to limit free expression (Cohen, 2013). Nevertheless, the Taft Court used the clear and present danger test in *Gitlow* to find that New York could prohibit the advocation of the violent overthrow of the government because such an advocation threatened the existence of the government (Holmes dissented here in similar fashion to his dissent in *Abrams*).

The Stone Court further narrowed the breadth of free expression in *Chaplinsky v. New Hampshire* (1942) where it decided that Chaplinsky's statements (calling the town marshal a "God-damned racketeer" and a "damned Fascist") were "fighting words" that were not protected under the First Amendment. The overarching case made by the Court was that the fighting words did not meet the goal of disseminating ideas to the public that the First Amendment was intended to protect, making New Hampshire's goal of maintaining the public peace more important. The first major instance in which the Court shielded an individual from indictment on constitutional free expression grounds was the Stone Court's decision in *West Virginia State Board of Education v. Barnette* (1943) where it ruled the mandating of a salute to the flag for schoolchildren to be unconstitutional, arguing that the imposition of a viewpoint is both an improper and ineffective means of achieving consensus.

The Warren Court established a new test of incidental restrictions on free expression in *United States v. O'Brien* (1968). The Court, facing the question of whether the burning of one's draft card during the Vietnam War was constitutionally protected, determined that the government may incidentally restrict free expression if the government's regulation furthers a substantive governmental interest, is unrelated to free expression, and is no more burdensome than required to achieve the substantive interest. In *Tinker v. Des Moines Independent School District* (1969), the Warren Court determined that students do not forfeit their First Amendment

rights upon entering the school building and that schools must demonstrate material and substantial harm to their operations in order to justify restrictions on free expression. In *Brandenburg v. Ohio* (1969), the Warren Court further refined the clear and present danger test established in *Schenck*; the Court's *per curiam* decision asserted that speech can only be restricted if it is "directed at inciting or producing imminent lawless action" and it is "likely to incite or produce such action."

The Burger Court held in *Board of Education, Island Trees Union Free School District No. 26 v. Pico by Pico* (1982) that libraries cannot restrict access to books which they find reprehensible due to them being centers for voluntary inquiry and the dissemination of information and ideas, reaffirming the Court's assertion in *Barnette* that the imposition of orthodoxy is improper. The Burger Court also made a distinction between political speech and other forms of speech, necessitating differing protections, in *Bethel School District No. 403 v. Fraser* (1986) where they decided that schools were allowed to restrict lewd speech in service of protecting educational interests.

In a decision that would likely turn heads today, the Rehnquist Court unanimously overturned the Minnesota Supreme Court's reversal of a trial court's dismissal of charges against teenagers who allegedly burned a cross on a black family's lawn in *R.A.V. v. City of St. Paul* (1992). Justice Scalia's opinion for the Court asserted that the ordinance was facially invalid because it would justify the prosecution of individuals for certain actions based upon the viewpoint being expressed, allowing other individuals to be free of potential indictment under the ordinance despite committing the same act (May, 2009). The Rehnquist Court also struck down certain provisions of the Communications Decency Act of 1996 (which sought to prevent minors from accessing obscene and/or indecent material on the internet) in *Reno v. ACLU* (1997) for a variety

of reasons, including a lack of a clear definition of "indecent" communications, no narrow tailoring of the legislation to prevent only minors from accessing certain content, and a lack of proof that the prohibited content did not present some social value. Furthering this, the Rehnquist Court decided in *Ashcroft v. ACLU* (2004) that the Child Online Protection Act (intended to prevent minors from accessing content that did not meet community standards) was overly prohibitive of First Amendment free expression because it would also prevent adults from accessing content that they had a right to access.

Many of the aforementioned cases worked with the assumption that the speech being protected was either an opinion or potentially represented some form of truth that may be valuable to public discourse, but the Court has also determined that plain falsehoods can also be granted protection under the First Amendment. The Roberts Court decided in *United States v. Alvarez (2012)* that the Stolen Valor Act (which criminalized the false claiming of receipt of certain military decorations) was unconstitutional because it allowed restrictions on speech that could cause no harm. Despite the claims being false, the Court asserted that they cannot be restricted except when the false claims could be demonstrated to cause harm. In response to suggestions by Justices, a revised version of the Stolen Valor Act was passed in 2013, this time specifying that the false claims must be made with the intent to receive something of value (Lilley, 2017).

Looking at how the understanding of First Amendment free expression has evolved in the legal system demonstrates key points that are of note. The first of these points is that the bounds of free expression have changed over time. As circumstances have changed, so too have attitudes towards free expression (even those of Supreme Court Justices) changed. Importantly, the trend has been towards protecting speech more than it had been protected in the past. Prior to *Gitlow*,

free expression wasn't even guaranteed to be protected from interference by the states. The Court's refinement of the clear and present danger test to include the stipulation that speech can only be restricted when the aim of the speech is both intended and likely to induce imminent lawless action seems to cause many of its earlier decisions to fail to pass constitutional muster (particularly those decisions made by the Stone Court). The second major note is that the Court's employment of the imminent lawless action standard (among other things) demonstrates our legal system's acknowledgement that free expression cannot be an absolute right. There is a recognition that free expression must be weighed against competing societal aims (national security and the protection of minors being chief among those weighed by the Court thus far). Finally, deference has been shifted from the government to the people in cases where First Amendment free expression is being considered; increasingly, the Court's disposition has moved towards putting the onus on the government to demonstrate both the necessity of their proposed restrictions and the lack of social value in the speech their restrictions affect. Ultimately, consideration of propositions relating to First Amendment free expression requires serious scrutinizing of the goals of such propositions and how they either achieve or undermine social values, including, but not limited to, free expression.

Section II: Free Expression in the Internet Age

A: Social Media Platforms as the New Public Forums

Any discussion of the freedom of expression that ignored how the forums used for expression have themselves changed over time would be remiss. The Court itself has acknowledged how such forums have evolved. In *Reno v. ACLU (1997)*, the Court argued that the internet operated as "'vast democratic fora' whose content 'is as diverse as human thought'" (Gey, 1998). Indeed, it is important to note that the modern internet is not only *a* forum, but it is also by far the largest and most influential forum shaping—and shaped by—both social and political attitudes. Affirming the importance of this, Justice Kennedy's opinion, joined by Justice Ginsburg, in *Denver Area Educational Telecommunications Consortium v. FCC* (1996) pointed out that "Minds are not changed in streets and parks as they once were. To an increasing degree, the more significant interchanges of ideas and shaping of public consciousness occur in mass and electronic media" (though this opinion was concurring in part and dissenting in part, Justice Kennedy gave a scathing critique of the plurality's refusal to consider how technological changes influenced First Amendment questions). The importance of electronic communication in the political process cannot be overstated. The rise of online information as the dominant source of news has empowered many to participate in modern politics by giving them information that they would otherwise not be able to access (Zhuravskaya, Petrova, & Enikolopov, 2020).

An important result of this spread of information has been decreasing support for incumbent regimes in a variety of countries with authoritarian and autocratic tendencies (Zhuravskaya et al., 2020). As an example, correlative analysis of Twitter conversations/hashtags with protests during the Arab Spring demonstrated a significant correlation between political coordination on the platform and protests in the following days (Steinert-Threlkeld, Mocanu, Vespignani, &

Fowler, 2015). Indeed, social media in general has been viewed as an integral part in the regime changes that occurred during the Arab Spring because of its ability to allow coordination between individuals who would otherwise have struggled to work in sufficiently large groups (Alhindi, Talha, & Sulong, 2012). A similar effect was found for the Russian social media site VK where coordination on the platform correlated with protests against the Russian government following the 2011 parliamentary election (Enikolopov & Petrova, 2015). Perhaps the most telling fact regarding the relationship between internet usage and citizen discontent is the widespread internet censorship that occurs in non-democratic states; the governments of those states clearly perceive public access to online information and communication channels as a threat to their existence (Zhuravskaya et al., 2020). Social media can thus be considered an important tool against oppressive governments.

The benefits of social media in the political realm are not isolated to non-democratic countries, however. Looking at the 2010 Congressional election in the United States, it was found that seeing posts about voting on Facebook increased the likelihood that someone would vote themselves; this effect was greater when individuals saw their Facebook friends post about having voted (R. M. Bond et al., 2012). A similar experiment was conducted again for the 2012 presidential election with the hypothesis that the effect would be minimal for a more high-stakes election where information from sources other than social media would be more prevalent; however, the same result was found (Jones, Bond, Bakshy, Eckles, & Fowler, 2017). Benefits are also extended to politicians in democratic nations. Primarily, they are able to interact more readily with their constituents as well as those whom they do not directly represent (Zhuravskaya et al., 2020). Beyond this, simply opening a Twitter account correlates with increasing campaign contributions, especially for non-incumbents (Petrova, Sen, & Yildirim, 2020).

Overall, the rise of new modes of communication and information dispersal have ushered in an unprecedented era of political awareness and activism that has been responsible for increasing the accountability of governments, whether that be to the benefit or detriment of those governments. Traditional sources of information (face-to-face communication, newspapers, books, leaflets, etc.), while still a part of spreading important political information, are now secondary, or supplementary, to online sources of information. The free expression exercised by individuals is no longer confined to verbal or printed material but is increasingly found in digital formats. While there are various benefits to this shift, the changing of the communication landscape has (unsurprisingly) not been without criticism. Before addressing those criticisms, however, it is important to frame the discussion in the context surrounding them. That context begins with the Communications Decency Act of 1996.

B: Communications Decency Act of 1996

In the early days of the internet, the New York Supreme Court took up the case of *Stratton Oakmont, Inc. v. Prodigy Services Co.* (1995). The plaintiff, a brokerage house, sought defamation charges against the online platform, Prodigy, a user of which had posted defamatory statements about the plaintiff (Siderits, 1996). The central question the court had to answer was whether Prodigy was acting as a publisher of the content in question or as a distributor of it (Siderits, 1996). The importance of this question laid in the fact that publishers and distributors of content had different liability for the content being published or distributed; publishers were considered liable for illegal content while distributors (such as a bookstore) were not (Johnson & Castro, 2021). As a result of Prodigy moderating the content on its site (as well as having compared itself publicly to a newspaper), the New York Supreme Court sided with Stratton Oakmont, deciding that Prodigy should be considered liable for the content posted by its users

since it was a publisher of its content (Siderits, 1996). To many today, this decision would seem outrageous because defamatory content is posted with great frequency on social media sites without an endless number of subsequent suits. This discrepancy can be accounted for by section 230 of the Communications Decency Act of 1996.

The Communications Decency Act of 1996's main objectives (as the name implies) were to prevent "obscene or indecent" material from being transmitted to children on the internet and to prevent anyone from knowingly displaying "patently offensive" material in a manner "available" to children there (Zeigler, 2009). As mentioned previously, a majority of the act was struck down as unconstitutional by the United States Supreme Court in *Reno*. One piece that was left standing, however, was section 230. Now heralded as one of the most important pieces of internet legislation, section 230 begins by proclaiming 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" 47 U.S.C. § 230 (1996). Section 230 also specifically protects moderation practices by including that no provider or user of an interactive computer service can be held liable for restricting access to material "that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected" 47 U.S.C. § 230 (1996). The ultimate goal of section 230 was to allow sites to prevent the proliferation of harmful or illegal content on their pages without punishing them for doing so with a ceaseless stream of suits (Johnson & Castro, 2021). Armed with protection from such suits, modern social media sites were able to come into existence (Johnson & Castro, 2021; Khanna, 2013). Despite section 230's role in allowing such companies to not only exist, but also flourish, many now have begun to paint it as a target for the reform of "big tech" (Khanna, 2013).

C: Reasons for the Calls to Alter or Abolish Section 230

The purported problems with "big tech"—a term that is ambiguous but seems to be at the forefront of the public conscience at the moment—are extraordinarily varied and, oddly, contradictory. People on both sides of the aisle have called for reform or outright repeal of section 230, but often for starkly different reasons; for instance, President Trump vetoed the National Defense Authorization Act in a bid to force Congress to repeal section 230 due to concerns about the censorship of conservatives while then President-Elect Joe Biden indicated that he also was in favor on repeal due to concerns that social media platforms weren't *censoring enough* content (Hamilton, 2021; Pietsch, 2020). These opposing reasons for the same legislative action represent a stark rift in the views held about section 230 and the effects that its repeal or reform would have. Repeal of section 230 simply cannot solve both problems that former President Trump and President Biden wish to rectify. Such a rift necessitates a careful review of the perceived problems with the operation of tech giants in order to start legislative considerations with a solid understanding of the present circumstances, especially considering the profound implications of altering section 230. The chief talking points of those calling for reform to section 230 are the spread of misinformation/disinformation (especially during elections and the COVID-19 pandemic), polarization aided by the algorithmic amplification of ideological segregation, and the deplatforming of speakers (especially political figures) (Reardon, 2021).

i. Misinformation and Disinformation

One of the potential problems with shielding social media platforms from liability—pointed out largely by Democratic lawmakers—is the spread of misinformation and disinformation (Pietsch, 2020). The thought is that social media platforms make it significantly easier for "fake

news" to spread rapidly for the same reason that real news is able to spread more rapidly: the nature of internet communication—traditionally viewed as one of its chief benefits—allows for extremely rapid transference of information between geographically distant locations. This thought, if true, is certainly threatening to social and political processes as it would substantially disrupt the effectiveness of public discourse on any issue. The arguments using this potential issue have largely focused on two areas of recent interest: elections and the COVID-19 pandemic.

The 2016 Presidential election saw the first widespread claims of election tampering and disinformation claims related to the cyber realm, following years of people of varied political leanings warning that such things would eventually occur (Keen, 2016). One of the chief worries cited—including by members of the United States' intelligence community—was the spreading of false information by twitter bots, especially bots manufactured by Russian actors (O'Connor & Schneider, 2017). It is reasonable to think that a barrage of false information planted into public discourse could alter voter perception of candidates, thus altering the outcome of the election. Fortunately, online platforms like Twitter and Facebook heard the criticisms of their tracking of inauthentic users, leading to significantly better measures against such accounts for the 2020 Presidential election (Howard, 2020).

Perhaps more importantly, however, is recognizing that the issue of fake news is far less widespread—and, in fact, far more concentrated—than previously thought. A study of fake news in the 2016 election showed that 80% of exposure to fake or extremely biased articles on Twitter was concentrated to just 1% of users (Grinberg, Joseph, Friedland, Swire-Thompson, & Lazer, 2019). Interestingly, the same study also found that 80% of shares of fake news came from only 0.1% of users (Grinberg et al., 2019). Overall, they concluded that the vast majority of users

obtained their information from mainstream outlets with a presence on Twitter while a very small subset of users (particularly, extreme right-wing and older users) were obtaining their information from sources distributing fake news (Grinberg et al., 2019). Similar findings were reported for Facebook during the 2016 election; shares of and interactions with fake news were sparse, with those over the age of 65 (controlling for ideology and partisanship) contributing seven times more to the shares than the youngest age group (Guess, Nagler, & Tucker, 2019).

The authors admit that the results were entirely unexpected; the traditional assumption is that ideology and partisanship play larger roles than demographic categories in the spreading of information (Guess et al., 2019). The main factor that seems to be at play is digital media literacy, something that the present data show is necessary for navigating new forms of media but is also drastically lower for older populations because the digital medium is an adjustment for them whereas it is the norm for younger populations (Livingstone, 2004; Neves & Amaro, 2012; Schäffer, 2007). Digital media literacy not only includes the basics of navigating the internet and the platforms therein, but also extends to abilities such as recognizing the differences between a real user and a bot as well as recognizing illegitimate information. These findings suggest not only that the effect of fake news on elections is drastically lower than typically suggested, but also that the present attempts at preventing the influence of fake news in that area (altering section 230, for instance) would likely not be effective. If we wish to combat fake news, it seems that the best course of action is working to increase digital media literacy. Importantly, this information also suggests that the problems fake news present may wane over time as digital media literacy (especially the identification of fake news) increases (in part due to the “aging out” of those with low digital media literacy and “aging in” of those who have grown up in the online communication age).

Perhaps more pressing than even the suggestion of insidious actions in our elections due to social media platforms is the possible role of social media misinformation and disinformation during the COVID-19 pandemic. A variety of scientific publications have pointed towards social media platforms as amplifiers of misinformation that has stifled national and global responses to the pandemic (Barua, Barua, Aktar, Kabir, & Li, 2020; Himelein-Wachowiak et al., 2021; Loomba, de Figueiredo, Piatek, de Graaf, & Larson, 2021; Tagliabue, Galassi, & Mariani, 2020). Importantly, it appears that the explanation of digital media literacy that seemingly accounts for the spread of misinformation related to elections does not account for the spread of misinformation relating to the pandemic; the clearest indication of this is the extraordinarily high vaccination rate (92.9% having received at least one dose for those between the ages of 65 and 74) of those with low digital media literacy in comparison to those that would have higher digital media literacy (Mayo, 2021). Thus, there seems to be some other factor at work during the current public health crisis that needs to be accounted for.

The present data do suggest something in this regard: digital media literacy is less able to combat false information when the information being digested requires a certain level of expertise to properly evaluate, especially when claims are competing. While individuals may be able to determine that a particular article isn't credible or that a particular online user is a bot, these abilities do not affect the ability to evaluate complex information coming from generally credible sources. Being a rapidly evolving situation, traditional media outlets, scientific publications, and health professionals have given good faith input to the situation that gets contradicted swiftly by new information (Tagliabue et al., 2020). This is further complicated by credible sources giving their input despite not having the expertise required to properly evaluate all of the information themselves, let alone distributing their analysis to others. Medical

professionals, for instance, may feel inclined to speak to the public about the ongoing situation despite not being particularly well-versed in virology or epidemiology (Tagliabue et al., 2020). The same issue arises for large media outlets (CNN, MSNBC, CNBC, FOX News, etc.). A thought commonly espoused to try to account for people believing and circulating false information about the pandemic is a distrust in medical experts; however, the data show that trust in medical experts does not predict the robustness of one's response to misinformation related to the pandemic (Loomba et al., 2021). Tying this in with the good-faith—but incorrect—statements made by health experts, it seems that the circulation of false information during the pandemic can largely be attributed to difficulties in parsing out good scientific information from poor scientific information (Loomba et al., 2021). This thought is strengthened by the fact that the spread of misinformation and reliable information about the pandemic is rather similar (Cinelli et al., 2020).

The difficulty in dealing with scientific information about the pandemic has become particularly problematic with the development of vaccines against COVID-19. There have been a variety of false claims circulated about the development, efficacy, and potential negative effects of the COVID-19 vaccines (Susarla, 2021). Importantly, this misinformation has been shown to decrease willingness to get vaccinated (Loomba et al., 2021). While social media platforms have allowed users to transmit false information about vaccines, blame cannot be laid entirely on them. Indeed, a major error in the push to reach herd immunity through vaccination has been poor communication of the science by those in the scientific community (Pollard & Bijker, 2021). Most false claims about the COVID-19 vaccines have started out with truths found in the data collected about them (Brumfiel, 2021). The critical point that has largely been handled improperly by those with the expertise to analyze such data is the extension of those kernels of

truth into larger, unsubstantiated claims (such as the vaccines potentially causing infertility) (Brumfiel, 2021). The basis of many of these claims in some truth makes it far more difficult for social media platforms to decide what information to leave up and what to remove.

Overall, social media itself does not appear to be the major cause of the spread of misinformation during the pandemic. While it is certainly a good medium for spreading misinformation, it also works to spread true information that is important for public health (constantly changing masking and social distancing recommendations, new data on disease incidence and community spread, updates on treatment and vaccine development, etc.). It also appears that cultural attitudes and predispositions have contributed to misinformation. Religiosity and conspiratorial thinking, for instance, negatively correlate with individual responses (following guidelines and recommendations) to COVID-19 (Barua et al., 2020). Furthermore, a study of the circulation of false pandemic-related information by Canadian Twitter users found that the majority of false information shared originated from US-based accounts and media outlets, pointing towards national coordination of information and responses as the major culprit rather than social media itself (Bridgman et al., 2021).

Underscoring the above discussion, it is important to note that very little work has evaluated differences in the spreading of false information between online and offline communication. Even if false information being spread online is problematic, the question remains whether that spread is significantly worse than the spread that would occur absent the internet (Zhuravskaya et al., 2020). Without internet communication, information could be gleaned from traditional sources such as books, newspapers, and television; however, the ideological leanings present in those mediums present their own problems in terms of spreading incorrect—or even manipulated—information. Beyond that, the spreading of information between people in the

relatively homogenous groups that individuals self-segregate into for day-to-day interactions could also readily spread and ingrain false beliefs to an extent at least comparable to that of the spreading of fake news on the internet (Zhuravskaya et al., 2020). This being the case, proposals for legislating away fake news seem not only ineffective, but also unnecessary given that political and social functions operated rather well with the problems presented by non-digital modes of communicating information. None of this is to say, however, that nothing need be done to combat the spread of false information on social media platforms. While the platforms themselves may not be the *reason* for misinformation being spread, their ability to *contribute to that spread* makes targeting them a reasonable strategy, especially for helping to prevent the disastrous response to the COVID-19 pandemic from occurring again in the future. Potential regulations in this area will be discussed in Section III.

ii. Polarization and Algorithms

Somewhat tied to the issue of misinformation is the thought that online communication lends itself to segmentation, allowing users to seek out and contain themselves within communities that agree with their ideology and presuppositions. Beyond just being able to actively seek out such communities, many argue that the algorithms used by social media companies to promote content similar to the content users appear to be interested in also serve to contain users within a sphere of ideas that they already hold (since the algorithm would promote content similar to what the users already seek out). The ultimate result, critics claim, is an increase in polarization. Two questions are thus raised. First, what is the extent to which this polarization occurs? Second, how does that polarization compare to the polarization that is caused by offline communication?

For the first question, the answer seems to be what critics contend: there is a high degree of ideological segregation in online communication networks (Cinelli, De Francisci Morales,

Galeazzi, Quattrociocchi, & Starnini, 2021). Interestingly, however, the degree of segregation can vary drastically between different online platforms; Facebook has a higher degree of homophily than Reddit, for instance (Cinelli et al., 2021). This distinction seemed to result from the option given on Reddit, but not Facebook, for users to tweak the settings of what appears in their feed (meaning the algorithm used by the site can be adjusted by the user). This leaves open the possibility that the abandonment of algorithms for displaying content is not necessary; perhaps simply altering the algorithms used (or allowing users to do so themselves) could resolve any issues seen regarding polarization.

Finding that individuals segregate ideologically on online platforms is not surprising, however. Indeed, it would be shocking to find the contrary. This isn't because people tend to think that online platforms are bastions of rigorous engagement with contrary views, rather, it is because we know that people tend to self-segregate regardless of the medium they are using to communicate and/or gather information. Decades of literature on the way people organize themselves into tight-knit groups and deal with information has shown that people, both online and offline, tend to organize into ideologically similar groups and seek out information that agrees with their predetermined views (confirmation bias) (DiPrete, Gelman, McCormick, Teitler, & Zheng, 2011; Nickerson, 1998; Peters, 2020). While algorithms do contribute to ideological segregation on online platforms, the natural tendency towards homophily of users seems to be a much larger factor in generating that segregation (Bakshy, Messing, & Adamic, 2015). The important question, then, is the second one raised above: how does the ideological segregation (and resulting polarization) of online platforms compare to offline networks?

For this question, the critics of online platforms appear to be wrong (or, at least, they appear to have overestimated the problem). Empirical estimates show that the segregation of online and

offline news consumption are comparable, but that interactions offline (face-to-face interactions with family, friends, etc.) are *more segregated* than interactions online (Gentzkow & Shapiro, 2010). Additionally, there does not appear to be a trend towards increasing online segregation over time (Gentzkow & Shapiro, 2010). Confirming these findings, models correlating age groups (a useful measure for the likelihood of using social media sites) with the observed increases in polarization demonstrate that polarization has increased the most for the age groups least likely to use social media sites and the internet more generally (Boxell, Gentzkow, & Shapiro, 2017). A possible explanation for the data is that users of online communication platforms interact more often with their "weak" or "peripheral" ties (more distant acquaintances and family members) than they would absent such platforms (Barberá, 2014). Since interactions with these weaker ties are likely to produce interactions with more disparate viewpoints than users would encounter in their face-to-face interactions, the user may experience a sort of moderating effect. Interacting with those weak ties could also cause the platforms' algorithms to begin suggesting more of those disparate viewpoints as well. Interestingly, political tolerance of marginalized groups has increased over time (though the increase varies from group to group) with the proliferation of online communication, potentially suggesting the moderation of viewpoints by social media interactions as a means of increasing tolerance (Boch, 2020).

Overall, the data seem to indicate two things regarding polarization on online platforms. First, communication on online platforms is characterized by homophily. Second, in stark contrast to the claims espoused by many fighting online platforms, the homophily (and resulting polarization) of online platforms is not drastically higher than that of offline forms of communication (and may actually be lower). With that in mind, targeting social media platforms' algorithms for reform does not seem to be the most efficient means of combatting the

observed increases in polarization. However, this doesn't mean that adjustments to these algorithms couldn't be of some benefit to combatting polarization (as evidenced by the differences in ideological segregation correlated to the ability to adjust one's own feed preferences).

iii. Deplatforming

Despite the above explorations finding little validity in some criticisms levied against online platforms, unique problems have arisen. One such problem, in my view, is the deplatforming (removal) of individuals by online platforms. In recent news, former President Trump was banned from Facebook, Twitter, and a variety of other online platforms (Alba, Koeze, & Silver, 2021; Luca, 2021). Importantly, Twitter and Facebook have indicated that their decisions regarding his deplatforming will stand even if he were to decide to run for office again during the 2024 election (though Facebook says it will be reviewing his ban in 2023, meaning at least that account could be reinstated) (S. Bond, 2021; Messenger, 2021). Rep. Marjorie Taylor Greene (R-Georgia) has also received multiple bans from Twitter, most recently receiving a one-week ban (Suciu, 2021). As of August 18th, 2021, five federal elected officials and one state representative have received bans on Twitter, Facebook, and/or YouTube (Ballotpedia, 2021). Bans of government representatives have only started to become an issue, but bans have long been levied against non-government users and are incredibly common (Luca, 2021). Twitter, for instance, banned 925,000 accounts for rules violations in the first half of 2020 alone (Luca, 2021).

While some have cheered the banning of individuals whom they view to be dangerous, others have pointed out that such bans, especially for political figures, serve to stifle important discourse and, perhaps more concerning, set dangerous precedents for the policing of political speech by tech giants (Morrison, 2021). There are a few issues with these bans that warrant

attention. First, such bans run contrary to basic principles of free expression. Considering that online platforms are the most important forums for social and political discourse in the modern era, bans from those platforms are the ultimate means of silencing speech. Figures may be able to try to voice their views in other ways, but no alternatives can adequately substitute for the largest online platforms. This is particularly important for political figures because social media platforms now serve as the primary means by which political figures are able to communicate with their constituents. Conversely, these platforms are the primary means by which constituents are able to see their representatives, thus influencing criticism by constituents and giving them important information that could influence their future voting. This consideration applies to political candidates as well. Newcomers, in the age of social media, need to have a social media presence to run a campaign that stands a chance. The banning of individuals from social media platforms is analogous to restricting speech in parks and on public sidewalks in the pre-internet era; however, such bans are more stifling considering the drastically greater reach that social media platforms provide. As noted in the first section, free expression is meaningless if dissident expression is disallowed. Overall, a society that values free expression cannot coherently permit absolute restriction of anyone's ability to participate in the primary medium of public discourse.

The second issue that warrants attention is the efficacy of banning figures from social media platforms. Even if it is granted that banning certain figures is desirable, the question of the effects of such bans on the spread of opinions/statements espoused by those figures remains. It seems that banning figures does not affect public discourse as intended. President Trump's ban, for instance, has not stopped his views on current topics from being expressed on social media platforms by others in his stead (Alba et al., 2021). Even outside of prominent figures being able to insert their views into public discourse post-banning, backfire dynamics have been observed

even before social media platforms were the dominant form of communication (Jansen & Martin, 2004). There is a general tendency, largely because of negative perceptions regarding censorship's opposition to free expression, for censored speech to become more sought out (Jansen & Martin, 2004). There are numerous examples of this sort of backfire. These dynamics are sometimes referred to as the "Streisand effect" because of the famous incident where Barbara Streisand tried to remove photos that had been taken of her and her home from online sites but ended up causing a drastically increased spread of those photos (Jansen & Martin, 2004). There are more pertinent examples for the discussion of political censorship backfiring as well. The banning of Alex Jones (an extreme right-wing political and social commentator) from social media platforms, for instance, amplified his voice through his own media outlet, *InfoWars* (Cofnas, 2019).

Third, censorship introduces a host of new problems that make the information being censored *more* dangerous. The first danger of deplatforming is forcing individuals into *more insulated* communities. Robert Bowers (a member of the alt-right), for instance, turned to the lesser-known platform Gab after receiving bans on the more mainstream platforms. On Gab, however, the community viewpoint on extreme views such as creating an ethno-state were reversed; "trolls" on that platform were those who didn't believe that Jewish people were nefarious (Cofnas, 2019). Following his being thrust into a situation where he exclusively interacted with those who would amplify his viewpoints, Bowers went on to commit the deadliest attack on Jewish-Americans in the nation's history (Santanam, 2019). Parler, a social media platform flocked to by conservatives in response to censorship on larger platforms, created an echo-chamber that was involved in drumming up support for the January 6th, 2021 attack on the Capitol (Bajak, Guynn, & Thorson, 2021). Importantly, banning of users from online

platforms (and encouraging them to engage in more insulated communities) prevents their exposure to counter-speech that could change their minds (which exposure, as the Supreme Court noted in *Alvarez,* is the preferred constitutional remedy to speech we dislike) (Cofnas, 2019). While some would characterize those with extreme beliefs as being "beyond reason", such characterizations are not only incorrect, but also actively harmful. It is not uncommon for individuals to have simply been misguided; assuming that their mind cannot be changed ensures that it won't be. Banning individuals from the primary forums of public discussion prevents them from seeing their views challenged, casting any hope of moderation of their views to the wayside (Cofnas, 2019). And, as noted above, these individuals will end up communicating on platforms where the opposing viewpoint is not expressed, and their present beliefs are only amplified. The final problem arising from censorship is that, for many of the views we would like to see removed from public discourse, censorship only gives more evidence (in the minds of those who hold the censored view) to the thing we aim to censor: "if it's necessary to ban discussion of the mainstream narrative, something fishy must be going on" (Cofnas, 2019). Ultimately, while there are figures that espouse beliefs that we wholeheartedly despise, we should not cheer when their voices are removed from public discussion. That removal runs contrary to any liberal principle of free expression, doesn't achieve what we wish it would, and only serves to increase the danger of the speech we aim to stifle.

Section III: Reforming Social Media Platforms

While many of the claims levied against social media platforms to argue for their reform are not as problematic as many have claimed, some changes to the current state of affairs are warranted. The current hands-off approach to social media platforms is insufficient given the breadth of their power over modern public discourse. Thoughtful consideration of remedying the presently observed issues with such platforms is needed to ensure that today's primary forums of public discussion are not distorted by those who host that discussion nor allowed to operate to the detriment of other public interests (such as public health). While private companies are not commonly thought to hold obligations under the First Amendment (since we traditionally conceive of the Constitution applying only to government action), First Amendment protections can, and should, be extended to actions by private entities that exercise a great degree of control over First Amendment free expression just as the Court extended such protections for actions by states in *Gitlow*. Importantly, absolutist views of free expression and property rights cannot be allowed to hamper the balancing of societal values and goals as they relate to regulation of social media platforms; as always, our values must be weighed against each other when they come into competition. Such weighing is the endeavor undertaken in this section.

A: Abolishing Section 230 is Untenable

Touted as a good reform by members of both sides of the political spectrum, abolition of section 230 is at the forefront of public discussion about reforming social media platforms (Pietsch, 2020; Reardon, 2021). This proposal, however, would be severely counterproductive. One of the chief intentions of section 230 was to foster the growth of online platforms where users could discuss topics with one another, taking advantage of the unprecedented ability to communicate with others that the internet offered; that goal is as important now as it was in

1996. Modern social media platforms have thrived because of section 230 and allow for the greatest free exchange of ideas that has yet been seen, but revocation of section 230 threatens that free exchange (Romano, 2018). Without immunity from liability, social media platforms would be encouraged to censor anything they *perceived* to generate a threat of being sued (Kelley, 2020). That situation would reverse the increases in the communication of ideas produced by social media platforms over the past two decades. Indeed, such a circumstance has already been observed with the limited reforms made to section 230; FOSTA-SESTA, a combination of two laws passed by Congress making section 230 immunity not applicable to the advertising of prostitution, has prompted companies to overly restrict any content they view as potentially in violation of the laws (Kelley, 2020; Romano, 2018). While complete repeal of section 230 is untenable, there are other measures that could prove useful.

B: Social Media Platforms as Common Carriers

A particularly attractive option, in my view, is tying section 230 immunity to common carrier status for online platforms. A common carrier is "a person or a commercial enterprise that transports passengers or goods for a fee and establishes that their service is open to the general public" (Institute, 2021). Internet platforms largely operate as common carriers already (the "good" they transport being speech). Their main service is to simply provide a medium for the expression of others. Some, including Justice Clarence Thomas, have suggested that the government would be within its right to simply declare internet platforms as common carriers due to the nature of how they operate (Brandom, 2021; Volokh, 2021). However, I think a more elegant, less brute-force approach would be to allow companies to self-designate as a common carrier with the benefit of the common carrier designation being the affording of section 230 immunity. If they chose to not self-designate as such, then they would be held liable as a

publisher of the content on their platform and remain able to exercise a large degree of discretion over the content posted by users. As a common carrier, however, social media platforms would have to transmit the posts of all users so long as those posts do not violate federal law. It is also important to note that the obligations of transmitting information from and to users should only apply for users in the United States since U.S. regulations could come into conflict with regulations made by other nations (though I would advocate for other nations adopting similar regulations).

The first question to tackle with regards to this proposition is the basis upon which it is being proposed. The principal reason is to continue allowing social media platforms that promote the free and open exchange of ideas and information to continue to flourish alongside that exchange. Choosing to be designated as a common carrier would allow these platforms to continue to operate more or less as they have thus far but would also obligate them to not ban users or censor constitutionally protected speech. The chief benefit, then, is further solidifying these platforms as open forums for the exchange of ideas which is necessary for a properly functioning democracy. The benefits of social media as the preeminent means of communication (discussed in Section II) cannot be gleaned by those that are excluded from that communication. The reduction in censorship that would follow is beneficial for two reasons. First, it contributes further to ensuring unfettered debate over important social and political topics occurs without undue influence by outside actors (the platforms themselves). Second, it would reduce the negative consequences of both good-faith and bad-faith censorship of extreme ideas (discussed in Section II). While I wouldn't contend that the current state of affairs is as horrendous as has been claimed (save for the outright banning of individuals, especially political figures), common carrier status would

serve as a protective measure against potential pernicious influences by social media platforms going into the future, adding an important safeguard to the primary forum of public discussion.

A concern that has been raised about restricting the moderation ability of platforms (which would result from the proposal presented here since platforms presently do ban constitutionally protected speech currently) is that it would spread undesirable speech on the platforms (Somin, 2021). The first criticism that I would levy against such concerns is that they run contrary to promoting free expression in the modern context. As long as free expression has been fought for, there have been concerns that we neuter our ability to get rid of speech that the majority of people dislike. However, a majority consensus on what speech is undesirable does not provide sufficient justification for the censoring of that speech. Free expression does not exist if we censor speech (or tacitly support the censoring of speech) that we dislike. Social media platforms are the modern public forums; we need to support free expression on these platforms with the same tenacity that we have defended free expression in other forums, even if that means we must contend with speech that we find to be loathsome. Importantly, our ability to counter such speech is far greater now than it has ever been before. Individuals from every corner of the country—indeed, every corner of the world—can speak contrary to those opinions that we find to be objectionable. There is also the added ability to block users that we find to be harassing ourselves or others, an ability that wouldn't be possible for someone choosing to walk in a park or on a sidewalk where someone was exercising their free expression to ends that we disagree with. The second argument I would present against those who wish to see the current moderation practices continue to be used is that such practices only exacerbate the problems that the censorship is ostensibly targeted at resolving (as discussed in Section II); limiting such moderation helps to prevent such backfire from occurring. The final point I would raise is that

while we may approve of some of the current moderation practices, we may find ourselves disagreeing with new practices in the future. Social media platforms have increased their restrictions over time and their restrictions are often shrouded in mystery (Wilson & Land, 2021). While many of the decisions around restricting speech on these platforms are made in response to cultural trends, those cultural trends (as discussed above) do not themselves justify further restricting free expression. Hate speech regulations are the most pernicious by far. The defining lines of hate speech are incredibly ambiguous and often lack context (whether speech about "white people" should be restricted, for instance) (Wilson & Land, 2021). Such ambiguity could easily be leveraged by social media platforms to ban speech that we find to fall within reason, something that governments have done time and time again in the past to stifle speech (the inclusion of Black Lives Matter in hate speech, the application of incitement laws to Keystone XL pipeline protesters, and countless instances in non-democratic nations) (Wilson & Land, 2021). Overall, while we may applaud many of the moderation decisions made by social media platforms, such moderation is harmful to free expression, backfires in ways that produce more harm, and puts public discussion in a dangerous position going into the future. Common carrier status could aid in resolving these issues.

An important note is that common carrier status would not prevent all removal of content. Common carriers are within their rights to restrict illegal content. Speech that is aimed at inciting to riot, for instance, is illegal, meaning that social media platforms could restrict such speech. Speech that is likely to create an imminent danger is not protected (as the Court articulated in *Brandenburg*) and so could be restricted by common carriers. The caveat here is that common carrier status would entitle users to appeal in the courts when their speech is restricted by platforms (a situation that I find preferable to the private arbitration presently practiced). The

allowance of restrictions against demonstrably dangerous speech during a public health crisis could also be afforded provided that Congress passed legislation deeming such speech to be illegal (this would be similar to the balancing between free expression and public interests done in cases like *Schenck*, *Abrams*, *Debs*, and *Bethel School District No. 403*). Alternatively, Congress could write an exception to the common carriage obligations of social media companies for speech that endangered public health during a public health crisis. The effect would be the same. Such an exception is similar to legislation already proposed by Sen. Klobuchar and cosponsored by Sen. Luján that would remove section 230 immunity for the carriage of misinformation related to a declared public health emergency on social media platforms (Levine, 2021). While such regulations could be useful in combatting flagrantly false information that endangers public health, they wouldn't solve the issues we faced during the COVID-19 pandemic (other helpful approaches are discussed later in this Section).

The most direct objection to treating social media platforms as common carriers is that it would entail government regulation of a private entity (Somin, 2021). This objection ignores the necessity of balancing the rights of private companies with other social values. Insofar as social media companies exercise (or, at least, have the ability to exercise) a large degree of control over public discourse, regulating their control is a reasonable step to ensuring free expression is robust. This is not a new step either. The government has regulated the actions of private companies in the past in order to protect citizens. For instance, antitrust laws prevent private companies from becoming monopolies that would harm the market, and thus harm citizens (Wilson & Land, 2021). Indeed, there is ample precedent from the Supreme Court that would indicate common carrier status as a constitutional remedy to the curbing of free expression in the

case of social media platforms (Wilson & Land, 2021). The Court stated the following in *Associated Press v. United States* (1944):

> "That Amendment rests on the assumption that the widest possible dissemination of information from diverse and antagonistic sources is essential to the welfare of the public, that a free press is a condition of a free society. Surely a command that the government itself shall not impede the free flow of ideas does not afford nongovernmental combinations a refuge if they impose restraints upon that constitutionally guaranteed freedom. Freedom to publish means freedom for all, and not for some. Freedom to publish is guaranteed by the Constitution, but freedom to combine to keep others from publishing is not. Freedom of the press from governmental interference under the First Amendment does not sanction repression of that freedom by private interests."

More recently, Justice Breyer asserted that "Requiring someone to host another person's speech is often a perfectly legitimate thing for the Government to do", citing *PruneYard Shopping Center v. Roberts* (1980), *Turner Broadcasting System v. FCC* (1994), and *Rumsfeld v. FAIR* (2006) to support his contention (Wilson & Land, 2021). While declaring social media platforms as common carriers could fall within the government's constitutional purview, my proposal of giving companies the option to self-designate would face less Constitutional pressure as it would just be the revocation of protection already granted by Congress, contingent on choices made by companies. Further, such revocation, even if challenged, could fall within Congress' spending power (the spending being the relief from costs of suits afforded to companies) (Wilson & Land, 2021).

The question, then, is whether social media platforms would choose to self-designate as common carriers given the restrictions that would be imposed upon them if they did make that

choice. I contend that they have a great incentive to do so. The chief reason that I reach this conclusion is that a platform that censored a large swath of posts by users for fear of suit (as they have shown would be the response to section 230 revocation) would almost certainly turn away many of its users, thus substantially decreasing the advertising revenue of the platforms. One of the concerns that has been raised about these platforms having a great deal of control over public discourse is that they are driven by a profit motive; we can utilize that motive to the benefit of public discourse (Somin, 2021). The question, though, is what effect common carrier designation would have on the profits of these companies. I find it unlikely that the designation would hamper the profits of social media platforms to any significant degree (the contrary thought comes from the possibility that advertisers wouldn't want to associate with a platform that has speech they don't want to be associated with). There are a few reasons for this. First, social media advertisements are rarely ever associated with a particular piece of content, meaning the advertisers don't become associated with particular messages (which makes it seem unlikely that advertisers would stop advertising on a platform that has speech they disagree with). Second, social media platforms can easily choose to move advertisements away from content that the advertisers would disagree with (YouTube already does something similar where advertisements are not played on channels that advertisers do not want to be affiliated with). Finally, even upon choosing common carrier status, the largest social media platforms would still be the largest, giving great incentive to advertisers to continue advertising on those platforms (because their advertisements would have the greatest reach on those platforms).

C: Additional Measures by Social Media Companies

While common carrier status would provide some useful measures to protect free expression, there are also initiatives that social media platforms could undertake themselves to further

resolve issues that would not be entirely resolved by common carrier status. The first area that these platforms could continue to make progress in is reducing the spread of misinformation. While the spread of misinformation on social media platforms does not appear to be worse than the spread of misinformation in offline communication (at least, not to the degree often claimed), there are reasonable measures that social media platforms could utilize to further aid in reducing misinformation while still falling in the bounds of common carrier obligations. The first such measure would be to continue working on fact-checking systems that can present counter-speech to false speech. While the largest platforms have already instituted fact-checking systems, these systems do not, as of yet, always present information in a way that is conducive to countering false information (Drutman, 2020). Working on increasing the effectiveness of fact-checking systems could genuinely help combat false information because, despite various claims to the contrary, fact-checking does help to alter beliefs and even the willingness to share the false information being countered (Drutman, 2020; Henry, Shuravskaya, & Guriev, 2020).

Social media platforms can also contribute to reducing polarization on their sites (even if that polarization is not, in relative terms, a large issue). The first means, mentioned earlier, is that these platforms can simply allow users to adjust their own feed preferences (similarly to Reddit) (Cinelli et al., 2021). Another option, for sites that host both general user feeds and groups that can be joined, is regulation of such groups. The limits of such regulations are unknown because, as far as I am aware, there are no truly comparable examples in the present common carrier case law. I would contend that Facebook, for instance, would likely be within their rights, even as common carriers, to suspend groups that are dedicated to spreading harmful content if Facebook made it such that groups are created by Facebook itself upon request by users. While the creation of groups on Facebook is presently controlled by users, Facebook could alter this in order to give

them leverage against the insulation of users within groups dedicated to spreading harmful ideas (forcing users into the larger forum where their ideas could still be expressed but would be subject to more counter-speech). Even absent this, groups dedicated to spreading information harmful to public health could be banned if the public health exceptions discussed earlier were in place. For instance, groups dedicated to convincing people to not get vaccinated (or groups denying the existence of SARS-CoV-2 or viruses more broadly) could still be removed if public health exceptions were in place.

D: Summary of Suggestions

While there will always be problems with social media communication (and communication more broadly), the suggestions presented here could serve to reduce such problems. The recommendations presented can be summarized as follows:

- Social media platforms should be given the opportunity to self-designate as common carriers (with the obligations carried by that designation applying only to users in the United States).
- Common carrier status for social media platforms should confer section 230 immunity to those platforms (and not to platforms not self-designating as common carrier).
- Exceptions to common carrier obligations should be granted for exigent circumstances such as public health crises to facilitate better public responses to such circumstances.
- Social media platforms should continue strengthening fact-checking mechanisms to bolster counter-speech on their platforms.
- Social media platforms should allow users to adjust the algorithms used to present them content in order to reduce ideological segregation.

- Social media platforms should ensure (if possible) that they maintain the ability to regulate groups so that the spreading of harmful information in insulated groups is reduced.

Conclusion

The way in which public discourse over important issues is conducted has changed dramatically since free expression was solidified as a right guaranteed to citizens by the First Amendment. Concomitant with this change have been changes in public attitudes towards and governmental regulations of free expression. Such changes, however, have not yet met the need for updates given the degree to which the exercise of free expression has changed. Communication now occurs primarily online; our laws must reflect this fact. Robust protections for free expression must be implemented wherever free expression is threatened. In the internet age, the largest threat to free expression comes from a lack of protections against influence by internet platforms; thus, protections in this area must be sought out and instituted. Leveraging section 230 immunity to limit such influence presents a valuable means by which to erect these protections. Challenges will doubtless be made against such leveraging, but those challenges must contend with the threats to free expression that we are faced with. I find that shielding free expression is—as has been the case in the past—more valuable than the negative consequences presented by that shielding.

Acknowledgements

This work is, in a direct sense, mine. But in another sense, it is the culmination of numerous discussions and experiences that I have had. I foremost thank Ian McKnight, a friend that thankfully agreed to undergo the arduous task of making it all the way through this work to suggest edits despite my lack of brevity. I also am very grateful to Eden Cantrell, Thaddeus Martin, and Kenna Spurlock for putting up with my odd rants about this paper as well as giving me varying perspectives on the issues I aimed to address. I thank Dr. Herman Mays because his response to an inquiry of mine pointed me in some new directions that proved fruitful. Finally, I would be remiss if I did not thank Dr. Susan Gilpin for her kind words and encouraging me to continue writing despite my assertion that I had no need to write outside of some limited circumstances. While the arguments laid out in this paper are not necessarily reflective of the views of those I've thanked (and, in many cases, likely are contrary to their views), their contributions were invaluable to me.

References

Case Law

Abrams v. United States, 250 U.S. 616 (1919)

Ashcroft v. ACLU, 535 U.S. 564 (2002)

Associated Press v. United States, 326 U.S. 1 (1945)

Barron v. Baltimore, 32 U.S. (7 Pet.) 243 (1833)

Bethel School District No. 403 v. Fraser, 478 U.S. 675 (1986)

Board of Education, Island Trees Union Free School District No. 26 v. Pico by Pico, 457 U.S. 853 (1982)

Brandenburg v. Ohio, 395 U.S. 444 (1969)

Chaplinsky v. New Hampshire, 315 U.S. 568 (1942)

Debs v. United States, 249 U.S. 211 (1919)

Denver Area Educational Telecommunications Consortium, Inc. v. FCC, 518 US 727 (1996)

Gitlow v. New York, 268 U.S. 652 (1925)

PruneYard Shopping Center v. Robins, 447 U.S. 74 (1980)

R.A.V. v. City of St. Paul, 505 U.S. 377 (1992)

Reno v. ACLU, 521 U.S. 844 (1997)

Rumsfeld v. Forum for Academic and Institutional Rights (FAIR), Inc., 547 U.S. 47 (2006)

Schenck v. United States, 249 U.S. 47 (1919)

Stratton Oakmont, Inc. v. Prodigy Services Co., 23 Media L. Rep. 1794 (N.Y. Sup. Ct. 1995)

Tinker v. Des Moines Independent School District, 393 U.S. 503 (1969)

Turner Broadcasting System v. FCC, 512 U.S. 622 (1994)

United States v. Alvarez, 567 U.S. 709 (2012)

United States v. O'Brien, 391 U.S. 367 (1968)

West Virginia State Board of Education v. Barnette, 319 U.S. 624 (1943)

Citations

Alba, D., Koeze, E., & Silver, J. (2021). What Happened When Trump Was Banned on Social Media. Retrieved from https://www.nytimes.com/interactive/2021/06/07/technology/trump-social-media-ban.html

Alhindi, W., Talha, M., & Sulong, G. (2012). The Role of Modern Technology in Arab Spring. *Archives des sciences 65*.

Bajak, A., Guynn, J., & Thorson, M. (2021). When Trump started his speech before the Capitol riot, talk on Parler turned to civil war. Retrieved from https://www.usatoday.com/in-depth/news/2021/02/01/civil-war-during-trumps-pre-riot-speech-parler-talk-grew-darker/4297165001/

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science, 348*(6239), 1130-1132. doi:10.1126/science.aaa1160

Ballotpedia. (2021). Elected Officials Suspended or Banned from Social Media Platforms. Retrieved from https://ballotpedia.org/Elected_officials_suspended_or_banned_from_social_media_platforms

Barberá, P. (2014). How social media reduces mass political polarization. Evidence from Germany, Spain, and the US. *Job Market Paper, New York University, 46*.

Barua, Z., Barua, S., Aktar, S., Kabir, N., & Li, M. (2020). Effects of misinformation on COVID-19 individual responses and recommendations for resilience of disastrous

consequences of misinformation. *Progress in Disaster Science, 8*. doi:https://doi.org/10.1016/j.pdisas.2020.100119

Boch, A. (2020). Increasing American Political Tolerance: A Framework Excluding Hate Speech. *Socius, 6*. doi:10.1177/2378023120903959

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature, 489*(7415), 295-298. doi:10.1038/nature11421

Bond, S. (2021). Trump Suspended From Facebook for 2 Years. Retrieved from https://www.npr.org/2021/06/04/1003284948/trump-suspended-from-facebook-for-2-years

Boxell, L., Gentzkow, M., & Shapiro, J. M. (2017). Greater Internet use is not associated with faster growth in political polarization among US demographic groups. *Proceedings of the National Academy of Sciences, 114*(40), 10612-10617. doi:10.1073/pnas.1706588114

Brandom, R. (2021). Clarence Thomas really wants Congress to regulate Twitter moderation. Retrieved from https://www.theverge.com/2021/4/5/22368045/clarence-thomas-social-media-regulation-knight-v-trump

Bridgman, A., Merkley, E., Zhilin, O., Loewen, P. J., Owen, T., & Ruths, D. (2021). Infodemic Pathways: Evaluating the Role That Traditional and Social Media Play in Cross-National Information Transfer. *Frontiers in Political Science, 3*(20). doi:10.3389/fpos.2021.648646

Brumfiel, G. (2021). The Life Cycle of a COVID-19 Vaccine Lie. Retrieved from https://www.npr.org/sections/health-shots/2021/07/20/1016912079/the-life-cycle-of-a-covid-19-vaccine-lie

Chomsky, N. (1980). Some Elementary Comments on The Rights of Freedom of Expression. Retrieved from https://chomsky.info/19801011/

Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences, 118*(9). doi:10.1073/pnas.2023301118

Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C. M., Brugnoli, E., Schmidt, A. L., . . . Scala, A. (2020). The COVID-19 social media infodemic. *Scientific Reports, 10*(1). doi:10.1038/s41598-020-73510-5

Cofnas, N. (2019). Deplatforming Won't Work. Retrieved from https://quillette.com/2019/07/08/deplatforming-wont-work/

Cohen, A. (2013). The Most Powerful Dissent in American History. Retrieved from https://www.theatlantic.com/national/archive/2013/08/the-most-powerful-dissent-in-american-history/278503/

Cost, J. (2017). James Madison's Lesson on Free Speech. Retrieved from https://www.nationalreview.com/2017/09/james-madison-free-speech-rights-must-be-absolute-nearly/

DiPrete, T. A., Gelman, A., McCormick, T., Teitler, J., & Zheng, T. (2011). Segregation in social networks based on acquaintanceship and trust. *Ajs, 116*(4), 1234-1283. doi:10.1086/659100

Drutman, L. (2020). Fact-Checking Misinformation Can Work. But It Might Not Be Enough. Retrieved from https://fivethirtyeight.com/features/why-twitters-fact-check-of-trump-might-not-be-enough-to-combat-misinformation/

Enikolopov, R., & Petrova, M. (2015). Media Capture. *Handbook of Media Economics, 1*, 687-700.

Gentzkow, M., & Shapiro, J. M. (2010). Ideological Segregation Online and Offline. *National Buereau of Economic Research Working Paper Series, 15916*. Retrieved from https://www.nber.org/papers/w15916

Gey, S. G. (1998). *Reopening the Public Forum-From Sidewalks to Cyberspace*.

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 U.S. presidential election. *Science, 363*(6425), 374-378. doi:10.1126/science.aau2706

Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances, 5*(1). doi:10.1126/sciadv.aau4586

Hamilton, I. A. (2021). Here's what could happen to Section 230 – the internet law Donald Trump hates – now the Democrats have both Houses. Retrieved from https://www.businessinsider.com/future-of-section-230-democrats-both-houses-2021-1

Henry, E., Shuravskaya, E., & Guriev, S. (2020). Checking and Sharing Alt-Facts. *SSRN*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3597191

Himelein-Wachowiak, M., Giorgi, S., Devoto, A., Rahman, M., Ungar, L., Schwartz, H. A., . . . Curtis, B. (2021). Bots and Misinformation Spread on Social Media: Implications for COVID-19. *J Med Internet Res, 23*(5). doi:10.2196/26933

Howard, B. (2020). How the 2020 election war on bots and trolls differs from 2016. Retrieved from https://www.cnbc.com/2020/10/27/how-the-2020-election-war-on-bots-and-trolls-differs-from-2016.html

Institute, L. I. (2021). Common Carrier. Retrieved from https://www.law.cornell.edu/wex/common_carrier

Jansen, S. C., & Martin, B. (2004). Exposing and Opposing Censorship: Backfire Dynamics in Freedom-of-Speech Struggles. *Pacific Journalism Review, 10*(1), 29-45. Retrieved from https://documents.uow.edu.au/~bmartin/pubs/04pjr.html

Johnson, A., & Castro, D. (2021). Overview of Section 230: What It Is, Why It Was Created, and What It Has Achieved. Retrieved from https://itif.org/publications/2021/02/22/overview-section-230-what-it-why-it-was-created-and-what-it-has-achieved

Jones, J. J., Bond, R. M., Bakshy, E., Eckles, D., & Fowler, J. H. (2017). Social influence and political mobilization: Further evidence from a randomized experiment in the 2012 U.S. presidential election. *PLoS One, 12*(4). doi:10.1371/journal.pone.0173851

Keen, A. (2016). How the Internet Is Threatening Our Freedom. Retrieved from https://www.politico.com/magazine/story/2016/05/2016-election-internet-campaign-facts-digital-new-media-213899/

Kelley, J. (2020). Section 230 is Good, Actually. Retrieved from https://www.eff.org/deeplinks/2020/12/section-230-good-actually

Khanna, D. (2013). The Law that Gave Us the Modern Internet—and the Campaign to Kill It. Retrieved from https://www.theatlantic.com/business/archive/2013/09/the-law-that-gave-us-the-modern-internet-and-the-campaign-to-kill-it/279588/

Levine, A. S. (2021). Klobuchar targets vaccine misinformation with Section 230 bill. Retrieved from https://www.politico.com/news/2021/07/22/klobuchar-vaccine-misinformation-section-230-bill-500554

Lilley, K. (2017). With new stolen valor law, Pennsylvania is the latest state to target military fakers. Retrieved from https://www.militarytimes.com/2017/06/28/with-new-stolen-valor-law-pennsylvania-is-the-latest-state-to-target-military-fakers/

Livingstone, S. (2004). Media Literacy and the Challenge of New Information and Communication Technologies. *The Communication Review, 7*(1), 3-14. doi:10.1080/10714420490280152

Loomba, S., de Figueiredo, A., Piatek, S. J., de Graaf, K., & Larson, H. J. (2021). Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nature Human Behaviour, 5*(3), 337-348. doi:10.1038/s41562-021-01056-1

Luca, M. (2021). Social Media Bans Are Really, Actually, Shockingly Common. Retrieved from https://www.wired.com/story/opinion-social-media-bans-are-really-actually-shockingly-common/

May, D. (2009). R.A.V. v. St. Paul (1992). Retrieved from https://www.mtsu.edu/first-amendment/article/270/r-a-v-v-st-paul

Mayo. (2021). U.S. COVID-19 vaccine tracker: See your state's progress. Retrieved from https://www.mayoclinic.org/coronavirus-covid-19/vaccine-tracker

Messenger, H. (2021). Twitter to uphold permanent ban against Trump, even if he were to run for office again. Retrieved from https://www.nbcnews.com/business/business-news/twitter-uphold-permanent-ban-against-trump-even-if-he-were-n1257269

Mill, J. S. (1859). *On Liberty*. London: J.W. Parker and son.

Morrison, S. (2021). Facebook and Twitter made Special World Leader Rules for Trump. What Happens Now? . Retrieved from https://www.vox.com/recode/22233450/trump-twitter-facebook-ban-world-leader-rules-exception

Neves, B. B., & Amaro, F. (2012). Too Old For Technology? How The Elderly Of Lisbon Use And Perceive ICT. *The Journal of Community Informatics, 8*(1). Retrieved from https://openjournals.uwaterloo.ca/index.php/JoCI/article/view/3061

Nickerson, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology, 2*(2), 175-220. doi:10.1037/1089-2680.2.2.175

Nunziato, D. C. (2003). Freedom of Expression, Democratic Norms, and Internet Governance. *George Washinton Faculty Publications and Other Works, 52*.

O'Connor, G., & Schneider, A. (2017). How Russian Twitter Bots Pumped Out Fake News During The 2016 Election. Retrieved from https://www.npr.org/sections/alltechconsidered/2017/04/03/522503844/how-russian-twitter-bots-pumped-out-fake-news-during-the-2016-election

Peters, U. (2020). What Is the Function of Confirmation Bias? *Erkenntnis*. doi:10.1007/s10670-020-00252-1

Petrova, M., Sen, A., & Yildirim, P. (2020). Social Media and Political Contributions: The Impact of New Technology on Political Competition. *Management Science, 67*(5), 2997-3021. doi:10.1287/mnsc.2020.3740

Pietsch, B. (2020). Trump and Biden both want to revoke Section 230, but for different reasons. Retrieved from https://www.businessinsider.com/trump-biden-want-to-revoke-section-230-for-different-reasons-2020-5

Pollard, A. J., & Bijker, E. M. (2021). A guide to vaccinology: from basic principles to new developments. *Nature Reviews Immunology, 21*(2), 83-100. doi:10.1038/s41577-020-00479-7

Reardon, M. (2021). Section 230: How it shields social media, and why Congress wants changes. Retrieved from https://www.cnet.com/news/section-230-how-it-shields-social-media-and-why-congress-wants-changes/

Romano, A. (2018). A new law intended to curb sex trafficking threatens the future of the internet as we know it. Retrieved from https://www.vox.com/culture/2018/4/13/17172762/fosta-sesta-backpage-230-internet-freedom

Santanam, R. (2019). Pittsburgh Synagogue shooting suspect pleads not guilty. Retrieved from https://apnews.com/article/religion-crime-shootings-pittsburgh-hate-crimes-8dcd0b6b03624272af4ab1d4e8a48412

Schäffer, B. (2007). The Digital Literacy of Seniors. *Research in Comparative and International Education*, *2*(1), 29-42. doi:10.2304/rcie.2007.2.1.29

Siderits, M. C. (1996). Defamation in Cyberspace: Reconciling Cubby, Inc. v. CompuServe, Inc. and Stratton Oakmont v. Prodigy Services Co. *Marquette Law Review*, *79*(4). Retrieved from https://scholarship.law.marquette.edu/mulr/vol79/iss4/7

Somin, I. (2021). The Case Against Imposing Common Carrier Restrictions on Social Media Sites. Retrieved from https://reason.com/volokh/2021/07/08/the-case-against-imposing-common-carrier-restrictions-on-social-media-sites/#

Steinert-Threlkeld, Z. C., Mocanu, D., Vespignani, A., & Fowler, J. (2015). Online social networks and offline protest. *EPJ Data Science*, *4*(1). doi:10.1140/epjds/s13688-015-0056-y

Suciu, P. (2021). Marjorie Taylor Greene Banned From Twitter Again. Retrieved from https://www.forbes.com/sites/petersuciu/2021/08/10/marjorie-taylor-greene-banned-from-twitter-again/?sh=12dea8f5659f

Susarla, A. (2021). Big tech has a vaccine misinformation problem – here's what a social media expert recommends. Retrieved from https://theconversation.com/big-tech-has-a-vaccine-misinformation-problem-heres-what-a-social-media-expert-recommends-164987

Tagliabue, F., Galassi, L., & Mariani, P. (2020). The "Pandemic" of Disinformation in COVID-19. *SN comprehensive clinical medicine*, 1-3. doi:10.1007/s42399-020-00439-1

van Mill, D. (2021). Freedom of Speech. *Spring 2021 Edition*. Retrieved from https://plato.stanford.edu/entries/freedom-speech/#Aca

Volokh, E. (2021). Treating Social Media Platforms Like Common Carriers? Retrieved from https://www2.law.ucla.edu/volokh/carrier.pdf

Wilson, R. A., & Land, M. (2021). Hate Speech on Social Media: Content Moderation in Context. *University of Connecticut School of Law Faculty Articles and Papers, 52*(3).

Zeigler, S. L. (2009). Communications Decency Act of 1996 (1996). Retrieved from https://www.mtsu.edu/first-amendment/article/1070/communications-decency-act-of-1996

Zhuravskaya, E., Petrova, M., & Enikolopov, R. (2020). Political Effects of the Internet and Social Media. *Annual Review of Economics, 12*(1), 415-438. doi:10.1146/annurev-economics-081919-050239