PUBLIC HEALTH LAW'S DIGITAL FRONTIER: ADDICTIVE DESIGN, SECTION 230, AND THE FREEDOM OF SPEECH

Matthew B. Lawrence[*]

A new generation of claims argues that addictive design by social media companies has caused a national mental health crisis, and so seeks to join nascent state legislative efforts in making addictive design by technology companies public health law's next frontier. But the threshold, global objections of leading social media platforms (including Facebook, Instagram, Snapchat, Tik Tok, and YouTube) to pioneering addictive design tort lawsuits—*In re Social Media Adolescent Addiction Litigation* in federal court and the *Social Media Cases* in California—suggest that state authority to regulate addictive design (through litigation or otherwise) will depend on the resolution of a conflict between two regulatory paradigms: the public health regulatory paradigm and the internet regulatory paradigm. The public health paradigm prizes federalism, with states historically playing a lead role in safeguarding the public's health through law—including against unwitting exposure to addictive products. Under this paradigm states would be permitted to develop and implement legal responses to an emerging public health threat through their courts and legislatures, as they have done with alcohol, gambling, opioids, and tobacco. The internet paradigm, on the other hand, usually insists on a "hands off" approach to regulation online, with broad federal preemption under

[*] Associate Professor of Law, Emory University School of Law; Affiliate, Harvard Law School, Petrie-Flom Center for Health Law Policy, Bioethics, and Biotechnology. Thank you to Gaia Bernstein, Margot Kaminski, Kenton MacDowell, Michael Ulrich, and Alan Rozenshtein, and Sasha Volokh for helpful suggestions and comments. Thanks also to the anonymous peer reviewers for the Journal of Free Speech Law, whose suggestions and critiques were very helpful in improving the draft. The author has no conflicts of interest to declare.

section 230 of the Communications Decency Act and often-prohibitive constraints under the First Amendment.

In the pioneering cases, the platforms argue that the internet paradigm makes pending lawsuits asserting addictive design claims non-starters, regardless of their merits. On the section 230 and First Amendment legal theories they advance, states could not regulate content-related addictive design by providers of interactive computer services (including social media platforms and some online video game manufacturers), no matter the evidence and no matter how intentional, effective, or harmful to kids or adults. Not surprisingly, the plaintiffs offer alternative views that would permit broad state regulation of addictive design.

This Article argues that, even if courts are unpersuaded by the broadest arguments in favor of a public health approach to regulation of addictive design, they should nonetheless reject the platforms' efforts to make addictive design a public-health-law-free zone. The public health and internet paradigms can be reconciled as a policy matter because addictive design threatens both public health and innovation online. The public health and internet paradigms can also be reconciled as a legal matter because even strong theories of section 230 and the First Amendment, properly understood, leave states a safe harbor in which to regulate much addictive design. Addictive design claims allege platforms engage in what psychologists call "operant conditioning" by using content-neutral intermittent reinforcement and variable reward techniques associated with slot machines to foster compulsion in users. These techniques need not entail content moderation or "editorial expression"; indeed, such techniques are ordinarily hidden from users, who may never realize they have been conditioned by a provider. State regulation of such content-neutral platform activity is not insulated from state public health regulation even under broad theories of the reach of section 230 and the First Amendment. To make maximal use of this safe harbor, public health researchers studying the harms of addictive design, legislators devising tailored regulatory responses, and courts adjudicating novel addictive design claims should remain mindful of the value of separating content-based addictive design claims from conditioning-based claims made in advancing public health law's digital frontier.

Introduction

States are today advancing laws to address the public health problem of addictive design—the knowing or negligent design of a product to foster compulsion in

a user[1]—by technology companies, such as social media platforms and video game developers.[2] Meanwhile, pioneering lawsuits in federal and state court seek to follow in the footsteps of eye-opening tobacco and opioid cases by applying existing tort and consumer protection laws to make litigation central to this next frontier of public health law.[3] Yet the authority of states to regulate addictive design online (whether through new legislation or application of existing laws in litigation) is today uncertain. Federal preemption under Section 230 of the Communications Decency Act and constitutional protection for the freedom of speech can limit state authority to regulate online activity, and the extent to which these laws insulate addictive design from state regulation has not yet been resolved. Prior scholarship has pointed to the threat of legal challenge looming over state regulation of addictive design but, with actual litigation merely hypothetical, largely declined to develop specifics.[4]

Addictive design brings two overarching legal paradigms—and two sets of constitutional principles—into apparent conflict. On the one hand, it is a funda-

---

[1] For definitions of the terms "addiction" and "addictive design," see *infra* notes 27–31 and accompanying text.

[2] *See, e.g.*, Dan DiFilippo, *N.J. Legislators Propose Punishing Social Media Companies for Kids' Online Addiction*, N.J. MONITOR (Feb. 22, 2023, 6:58 AM), https://newjerseymonitor.com/2023/02/22/n-j-legislators-propose-punishing-social-media-companies-for-kids-online-addiction/ (describing addictive design legislation in California, Maryland, Minnesota, and New Jersey); *see also* Mary Clare Jalonick, *Congress Eyes New Rules For Tech, Social Media: What's Under Consideration*, KETV (May 8, 2023, 1:52 AM), https://www.ketv.com/article/whats-under-consideration-congress-eyes-new-rules-for-tech-social-media/43821405 (describing bills pending in the United States Senate, including a requirement that minors have the option to "disable addictive product features").

[3] *See In re* Soc. Media Adolescent Addiction/Pers. Inj. Prods. Liab. Litig., No. 4:22-MD-03047-YGR, 2023 WL 2414002, at *1 (N.D. Cal. Mar. 8, 2023); Soc. Media Cases, No. JCCP 5255, Lead Case No. 22STCV21355, 2023 WL 6847378 (Cal. Super. Ct. L.A. County Oct. 13, 2023).

[4] Kyle Langvardt, *Regulating Habit-Forming Technology*, 88 FORDHAM L. REV. 129, 175 (2019) (discussing Section 230); *id.* at 178 (discussing addictive design and the First Amendment); Luke Morgan, *Addiction and Expression*, 47 HASTINGS CONST. L.Q. 197, 201 (2020) (discussing First Amendment challenges); GAIA BERNSTEIN, UNWIRED: GAINING CONTROL OVER ADDICTIVE TECHNOLOGIES 78 (2023). For a discussion of substantive theories of liability for addictive design—some of which appear in the complaint in *In re Social Media Addiction*—that includes some discussion of interactions with Section 230, see Nancy S. Kim, *Beyond Section 230 Liability for Facebook*, 96 ST. JOHN'S L. REV. 353, 372 (2022).

mental tenet of public health law and federalism that the states, in signing the Constitution, did not relinquish their independent authority to safeguard the public health, and, so, that even constitutional limits on state authority may be overcome when states act to further the public's health.[5] On the other hand, it has become a fundamental tenet of regulation of the internet that states have regulatory authority only within limited (albeit evolving and under-defined) bounds created by Section 230 and the First Amendment.[6]

Can these two apparently conflicting regulatory paradigms be reconciled and, if not, which should trump? Is public health law's next frontier wide open for states to force platforms to disclose internal research and craft responses to mounting public and expert concerns about the reality and potential of addictive design, is it closed entirely, or is it something in between? This Article seeks to shed some light on these emerging questions and to highlight a middle-ground path forward employed by Judge Kuhl in an insightful early ruling in October 2023 in the *Social Media Cases*[7]—a path the Article labels the "neutrality triangulation" approach—that partially reconciles Section 230 and the First Amendment, on the one hand, with states' public health interest in protecting their residents (especially their kids) from unwitting exposure to addictive products, on the other. The Article does so in five parts.

Part I provides background. It begins by describing historical state public health regulation of addictive products, including gambling products like slot machines believed to contribute to behavioral addiction through "operant conditioning" techniques such as intermittent reinforcement and variable reward. It then describes mounting concerns about addictive design in the development of new

---

[5] Jacobson v. Massachusetts, 197 U.S. 11, 24–25 (1905) ("The authority of the state to enact [a vaccination requirement] is to be referred to what is commonly called the police power,—a power which the state did not surrender when becoming a member of the Union under the Constitution."); *see generally* WENDY E. PARMET, CONSTITUTIONAL CONTAGION (2023) (discussing COVID-era rulings, including some questioning state authority); Wendy E. Parmet, *Health Care and the Constitution: Public Health and the Role of the State in the Framing Era*, 20 HASTINGS CONST. L.Q. 267 (1993); *see also* Michael Ulrich, *A Public Health Law Path for Second Amendment Jurisprudence*, 71 HASTINGS L.J. 1053, 1070–78 (2020) (collecting sources).

[6] *See, e.g.*, Zeran v. Am. Online, Inc., 129 F.3d 327, 334 (4th Cir. 1997) (Section 230 makes clear that "Congress' desire to promote unfettered speech on the Internet must supersede conflicting common law causes of action").

[7] *Soc. Media Cases*, 2023 WL 6847378, at *30–35 (Section 230), *35–39 (First Amendment).

technologies before turning to the "nitty gritty" of *In re Social Media Adolescent Addiction*. *In re Social Media Adolescent Addiction* is the first major federal addictive design case; the *Social Media Cases* pending in California Superior Court represent the most advanced major state case. In the *In re Social Media Adolescent Addiction* case, a (growing) variety of plaintiffs, including school districts, individuals, and state and local governments, allege that several major platforms designed their products to foster compulsion in unwitting adolescents, with widespread and harmful results.[8]

At this writing, the platforms' motion to dismiss all claims *in toto* on Section 230 and First Amendment grounds has recently been denied in part and granted in part by Judge Rogers in *In re Social Media Adolescent Addiction* and Judge Kuhl has issued an opinion on an analogous motion in the *Social Media Cases*, so the parties' (and especially the platforms') positions have been staked out and the first judicial rulings have been rendered.[9] Parts II and III unpack the available legal theories developed in these early cases as I understand them.

Regarding Section 230, to a significant extent the platforms' arguments hinge on the legal question whether Section 230 preempts state regulation of platforms' content prioritization choices (including user-specific content recommendations)—a question considered but not resolved by the Supreme Court in *Gonzalez* (which I call the "matchmaking question" in reference to Judge Katzmann's lead-

---

[8] *See infra* Part I.C for a description of the plaintiffs' claims.

[9] Judge Rogers granted in part and denied in part the platforms' motion to dismiss on November 14, 2023, after this Article went to press. *See In re* Soc. Media Adolescent Addiction/Pers. Inj. Prods. Liab. Litig., No. 4:22-MD-03047-YGR, 2023 WL 7524912 (N.D. Cal. Nov. 14, 2023). This Article therefore does not fully unpack the substance of the order, addressing arguments as they are presented by the parties as indicia of how the legal issues will be presented in other courts and, potentially, on appeal. That said, Judge Rogers' opinion is largely consistent on the Section 230 and First Amendment issues with Judge Kuhl's approach in *Social Media Cases*, which I endorse here, and with the related "safe harbor" framework I describe. *E.g.*, *id.* at *18 (refusing to dismiss claims as to filter-related effects because such effects are "neutral, non-expressive tools"). For another recent decision related to digital public health rendered after the Article went to press, *see* Neville v. Snap, No. 22-STCV-33500, op. at 11 (Cal. Super. Ct. L.A. County Jan. 2, 2024) (wrongful death case related to facilitation of fentanyl sales could proceed; holding claim outside of "sweet spot" of section 230 for activity removing or declining to remove third party content).

ing opinion[10] articulating it). The platforms argue for complete federal preemption of addictive design claims through the combination of the legal premise that Section 230 does bar states from regulating all content moderation choices (despite the uncertainty surrounding the matchmaking question after *Gonzalez*) and the factual premise that addictive design claims challenge such choices. The plaintiffs, for their part, argue that Section 230 is inapplicable to platforms' upstream "design" choices, except insofar as those choices are about whether and how to censor content—an argument that also seems to implicate the *Gonzalez* question, albeit from a different angle.

Part II emphasizes that while the matchmaking question is certainly important, many addictive design claims would fall outside the reach of Section 230 even if that question were resolved in favor of preemption because addictive design can, but need not, involve content moderation. Many addictive design claims focus on platform conduct that is neutral to the content of user speech, such as the alleged use of intermittent reinforcement and variable reward to stimulate compulsive use.[11] Such claims do not challenge the specific subject matter of content presented or the extent to which a platform censors, prioritizes, or recommends particular content. To find such claims preempted, it would be necessary to read Section 230 to preempt not *only* state regulation of platforms' decisions discriminating among users' content, but *also* platforms' content-neutral choices. Section 230's text, history, purposes, and precedent indicate that Section 230 does not apply to state laws insofar as they regulate platform conduct that is content neutral as to users' expression (including platform choices regulating the time, place, and manner of users' expression in a non-discriminatory way).[12] Thus, this Article agrees with Judge Kuhl's use of this neutrality triangulation[13] approach to reject

---

[10] *See* Force v. Facebook, 934 F.3d 53 (2d Cir. 2019) (Katzmann, J., concurring in part and dissenting in part) (framing certain platform activities as not merely recommending or prioritizing, but "matchmaking" because of the degree of individual-specific, personal information involved).

[11] *See infra* Part IV (discussing extent to which addictive design claims are content based).

[12] *In re* Zoom Video Commc'ns, Inc. Priv. Litig., 525 F. Supp. 3d 1017 (N.D. Cal. 2021); *see infra* Part III.

[13] I use the term "neutrality triangulation" to describe this approach in order to emphasize that the key Section 230 question is not whether a state law is content neutral (facially or as applied) as between the state and the platform or provider, but rather whether a state law, facially or as applied, regulates platform conduct that is content neutral as between the platform and the provider. *See* Jack M. Balkin, *Free Speech Is a Triangle*, 118 COLUM. L. REV. 2011, 2014 (2018) (noting

platforms' legal objection to addictive design claims targeting their content-neutral activities.[14]

As for the First Amendment, Part III notes that whether regulation of addictive design implicates the freedom of speech at all is an open question that largely depends on the broader question whether platforms' (and other providers')[15] content moderation activities are expressive. This big-picture question has split the Fifth and Eleventh Circuits, and the Supreme Court has granted *certiorari* and is expected to resolve the issue this term.[16] But again, it would be a mistake to see state authority to regulate addictive design as rising and falling entirely with this larger legal dispute about state authority to regulate content moderation, because many addictive design claims challenge platform activity that is neutral to the content of users' expression, *i.e.*, that does not involve censoring, failing to censor, prioritizing, or recommending particular users' content at all. Thus, even on the Eleventh Circuit's expansive view of First Amendment protection for platforms and providers—on which their content moderation choices are a form of "editorial" expression, like the choices of a newspaper on what opinions to feature on its editorial page—addictive design claims challenging platforms' use of operant conditioning techniques would not implicate the freedom of speech. Moreover, the public health concerns underlying efforts to regulate addictive design—particularly concerns for protecting users from being unwittingly exposed to addictive products[17]—are content neutral, and so could support properly tailored regulation even of expressive (and so First Amendment-protected) platform activity.

---

triangular relationship between state, platform, and user); *see infra* Part II.C (describing triangulation approach).

[14] Soc. Media Cases, No. JCCP 5255, Lead Case No. 22STCV21355, 2023 WL 6847378, at *30–35 (Cal. Super. Ct. L.A. County Oct. 13, 2023) (Section 230); *id.* at *35–39 (First Amendment).

[15] Section 230 applies not just to platforms but to all interactive computer service providers. So does the discussion here, though my focus—like that of the pending cases I discuss—is platforms.

[16] *See infra* Part III.A.

[17] Part III.C discusses states' substantial interest, grounded in states' interest in safeguarding their residents' liberty, in protecting residents from unwitting exposure to addictive design.

Part IV offers illustrative applications based on the claims in the *In re Social Media Addiction* case. Even if state laws regulating addictive design (including claims to enforce generally applicable law) are preempted by Section 230 or limited by the First Amendment to the extent they regulate content-based platform conduct (larger legal questions the Article describes but does not attempt to resolve), states would still have authority to regulate addictive design through content-neutral platform conduct. It is therefore possible to identify allegedly addictive design techniques that states may regulate even under broad theories of Section 230 preemption and First Amendment coverage, such as certain platforms' alleged deliberate delay of notifications about "likes" received until a sizable "jackpot" has accumulated, or their failure to warn about the alleged tendency of their products to stimulate compulsive use.[18] A conclusion summarizes the Article's contribution.

Before proceeding, a caveat about the assertions in this Article: It is essential to the legal analysis here to describe widespread concerns voiced about addictive design as well as allegations and arguments in the *In re Social Media Adolescent Addiction Litigation* and the *Social Media Cases*. By doing so, I do not mean to assert that those concerns or allegations and arguments are, in fact, accurate—though, to be clear, I personally share concerns that addictive design is a threat to public health, especially for kids. As a court would do in resolving a motion to dismiss,[19] I assume the correctness of factual allegations made about addictive design in these cases in order to isolate the legal questions such facts give rise to. Adversarial litigation on the merits will offer an opportunity to test the factual validity of addictive design claims against platforms (and, through discovery, to balance the information asymmetry between platforms, on the one hand, and policymakers and the public, on the other, about how addictive design actually impacts users). As for the parties' arguments, it is also necessary for me to characterize, summarize, and extend their theories to new contexts, which brings the risk that I may have misunderstood them. I have done my best to characterize the parties' positions accurately. I try to provide pincites and quotes to support my character-

---

[18] *E.g.*, Plaintiff's Amended Master Complaint (Personal Injury) ¶ 79, *In re* Soc. Media Adolescent Addiction/Pers. Inj. Prods. Liab. Litig., No. 4:22-MD-03047 (N.D. Cal. Apr. 14, 2023) [hereinafter Compl.].

[19] Barnes v. Yahoo!, Inc., 570 F.3d 1096, 1098 (9th Cir. 2009) (the court must "accept as true the facts alleged in the complaint and construe them in the light most favorable to the plaintiff").

izations and I suggest the reader refer directly to the pleadings in each case for a comprehensive description of the parties' arguments.

I. A NEW FRONTIER FOR PUBLIC HEALTH LAW

A. Public Health Concerns Surrounding Addictive Design

In 2023, the entrepreneurs who design apps and websites have near-limitless control over the who, what, when, where, and how of their users' experiences, including, for "interactive computer service providers" (like social media platforms and some online video game makers), the who, what, when, where, and how of their users' communication with one another.[20] Many people believe that the dominant social media platforms have achieved their dominance by using this power over the who, what, when, where, and how of communication and user experience to construct worlds that look and work more like a casino than a marketplace of ideas.[21] According to a chorus of parents, kids, teachers, scholars, activists, and former platform employees, platforms have designed their apps (either knowingly or negligently) to foster compulsive use in unwitting users, including both kids and adults, with widespread and harmful effects on public health.[22]

---

[20] As just one example, Facetime lets me speak on video to my niece from 500 miles away as a chicken or cat.

[21] For illustrative comparisons between internet-based (especially smartphone) activities and slot machines, see, for example, Sheldon A. Evans, *Pandora's Loot Box*, 90 GEO. WASH. L. REV. 376, 378 (2022) ("Loot boxes are a mix between pulling a slot machine lever and buying a pack of trading cards."); Kim, *supra* note 4, at 375 ("Users get a dopamine rush with each new notification in a way that one former Facebook engineer compared to a gambler at a slot machine."); Langvardt, *supra* note 4, at 151 ("Speed reinforces the compulsive flow of a slot machine, and social apps thrive on the same phenomenon."); Catherine Price, *Trapped—the Secret Ways Social Media Is Built to Be Addictive (and What You Can Do to Fight Back)*, SCI. FOCUS (Oct. 29, 2018, 10:00 PM), https://www.sciencefocus.com/future-technology/trapped-the-secret-ways-social-media-is-built-to-be-addictive-and-what-you-can-do-to-fight-back (making such a comparison because both slot machines and social media employ intermittent reinforcement and variable reward); Vikram R. Bhargava & Manuel Velasquez, *Ethics of the Attention Economy: The Problem of Social Media Addiction*, 31 BUS. ETHICS Q. 321, 321 (2020) ("Social media companies commonly design their platforms in a way that renders them addictive.").

[22] *See infra* Part I.C (describing various examples cited in pending litigation); BERNSTEIN, *supra* note 4, at 78; Matthew B. Lawrence, *Addiction and Liberty*, 108 CORNELL L. REV. 259, pt. II.C, at 290–98 (2023).

Underlying these concerns is a phenomenon more narrow and intrusive than notions of "dark patterns" or "manipulation,"[23] which include nudges and choice architecture tricks that may steer a user during a particular interaction with a platform or device but do not follow the user over time after she turns off her device—whether she wants them to or not—into her car, home, or office.[24] This explains scholars' use of terms such as "habit-forming technology"[25] and "addictive design"[26] to refer to product design that intentionally or reasonably foreseeably increases the likelihood that users will develop a potentially harmful compulsion (*i.e.*, a persistent, intrusive urge) to use a product. Here I use the term "addictive design" to focus on design techniques that foster persistent, intrusive urges because whether understood medically or in ordinary parlance, compulsion is at the core of the term "addiction."[27]

---

[23] *See* Micah L. Berman, *Manipulative Marketing and the First Amendment*, 103 GEO. L.J. 497 (2015); Ryan Calo, *Digital Market Manipulation*, 82 GEO. WASH. L. REV. 995 (2014).

[24] *Cf.* Kyllo v. United States, 533 U.S. 27 (2001) (use of thermal imaging to see into home constitutes a search, even if done from public vantage point).

[25] *See, e.g.*, Langvardt, *supra* note 4.

[26] *See, e.g.*, Caleb N. Griffin, *Systemically Important Platforms*, 107 CORNELL L. REV. 445, 449 (2022) ("Understanding—and regulating—the addictive design at the core of so many Big Tech platforms is a necessary complement to work on Big Tech's antitrust, privacy, and speech issues.").

[27] The term "addiction" means different things from the standpoint of different linguistic communities. *See* NANCY D. CAMPBELL, DISCOVERING ADDICTION: THE SCIENCE AND POLITICS OF SUBSTANCE ABUSE RESEARCH 1–11 (2007) (discussing construction of term). When 56% of Americans self-describe themselves as "addicted" to their smartphones, Alex Kerai, *2023 Cell Phone Usage Statistics*, REVIEWS.ORG (July 21, 2023), https://www.reviews.org/mobile/cell-phone-addiction/, they do not seem to be using the term in precisely the same way as are medical professionals when they diagnose a person with an "addiction" sufficient to trigger disability protections, insurance coverage of treatment, and so on under the diagnostic standards established by the profession through the Diagnostic and Statistical Manual of Mental Disorders. *See* Lawrence, *Addiction and Liberty*, *supra* note 22, at 310–12 (discussing lay and medical definitions of addiction). There is, however, a core point of overlap among these differing understandings: As understood in medicine and in lay terminology, "addiction" entails a compulsion (persistent, intrusive urge) to engage in a harmful behavior. *Id.* Variation around this core understanding of addiction largely surrounds the question of what type and degree of "harm" is required and (relatedly) whether "harm" is assessed subjectively (from the internal perspective of an individual) or objectively (from the external perspective of a person viewing the individual). The medical profession employs a largely (and unavoidably) objective definition of "harm" that focuses on functional

Addictive design is believed to be particularly concerning in an "attention" economy in which users' time equals money and a network's power and attractiveness to users depends on how many users it can grab and keep.[28] Because network effects give a competitive advantage to the platforms with the *most users*, addictive design gives platforms that do it successfully an artificial competitive advantage that is less likely to be "fixed" by the market insofar as non-addictive competitors are at a disadvantage in accumulating the user engagement necessary for success.[29]

---

impairment, while lay definitions are more subjective and may count any unwanted impulse as sufficient. *Id.* at 312 nn.266–67 and accompanying text. Using the term addiction also connects emerging concerns about compulsive use of technology with longstanding epidemics of drug and alcohol addiction. I explain in *Addiction and Liberty* my view that, while we must be careful not to suggest a false equivalence, shying away from acknowledging commonalities between behavioral addiction and drug addiction may only further entrench societal stigma surrounding drug addiction. *Id.* at 335–42.

[28] Bhargava & Velasquez, *supra* note 21, at 337 ("Social media companies . . . are advancing their own ends when they get users to engage and remain engaged with their social media platforms."); Tim Wu, *Blind Spot: The Attention Economy and the Law*, 82 ANTITRUST L.J. 771, 783 (2019) ("[O]ur attentional decisions can be compared to other consumer decisions, such as spending money.").

[29] *See* Nitish Pahwa, *Swapping a Twitter Habit for a Threads One*, SLATE (July 31, 2023, 1:48 PM), https://slate.com/technology/2023/07/threads-features-meta-twitter-feed-addiction.html ("[T]he best strategy to keep Threads competitive is to embrace the key element that positioned Facebook, TikTok, Instagram, and Twitter (long before X) to become some of the world's most popular websites: the addiction factor."); Maya MacGuineas, *Capitalism's Addiction Problem*, THE ATLANTIC, Apr. 2020, https://www.theatlantic.com/magazine/archive/2020/04/capitalisms-addiction-problem/606769/ ("In a well-functioning market, consumers have the freedom to act in their own self-interest and to maximize their own well-being."); Shota Ichihashi & Byung-Cheol Kim, *Addictive Platforms*, 69 MGMT. SCI. 1127 (2022) (presenting economic model under which "if attention is scarce, increased competition reduces the quality of services because business stealing incentives induce platforms to increase addictiveness"); James Niels Rosenquist, Fiona M. Scott Morton & Samuel N. Weinstein, *Addictive Technology and Its Implications for Antitrust Enforcement*, 100 N.C. L. REV. 431, 484 (2022) ("because increased consumption of social media may simply reflect low quality and addiction, it need not increase consumer welfare"); *see also id.* ("[t]he assumption that more consumption of addictive digital products leads to increased utility is not justifiable based on the medical and economics literature").

Driven by these perverse competitive pressures, platforms are alleged to have designed their products to take advantage of "operant conditioning"[30] mechanisms developed in psychology and long associated with slot machines by tailoring the timing and predictability of rewards and interactions in ways that foster compulsion.[31] Platforms' designs are also alleged to highlight the most attention-grabbing or engaging content. (Specific allegations of ways that platforms foster compulsion in users are described in subpart C.)

Public health concerns surrounding addictive design are evident in the United States Surgeon General's recent "Social Media and Youth Mental Health" advisory. That advisory describes harms of, *inter alia*, "compulsive or uncontrollable use," and notes that cutting-edge studies of "people with frequent and problematic social media use" showed that such users "can experience changes in brain structure similar to changes seen in individuals with substance use or gambling addictions."[32] A growing body of research underlying the Surgeon General's advisory is seeking to understand all aspects of addictive design and its effects, from the direct user harms of compulsive use to second-hand harms from "distracted"

---

[30] The term "operant conditioning" is associated with the work of psychologist B. F. Skinner. B. F. Skinner, *Superstition in the Pigeon*, 38 J. EXPERIMENTAL PSYCH. 38, 168 (1948); B. F. SKINNER, THE BEHAVIOR OF ORGANISMS: AN EXPERIMENTAL ANALYSIS (1938). His experiments suggested that the incentive to obtain a reward could be cultivated into a compulsion to perform the rewarded activity through the use of variable timing and uncertainty, among other tactics. *See also* NITA FARAHANY, THE BATTLE FOR YOUR BRAIN 156–57 (2023) (discussing various approaches to technology development that "exploit[] shortcuts in our brains").

[31] For example, an "operant conditioning" mechanism featured in many slot machines is the "near miss." The theory behind the "operant conditioning" understanding of the "near miss" mechanism is that leading a user to believe they have nearly obtained a reward but just barely failed to do so (as when a slot machine stops one number shy of a reward, or a children's "claw" game lifts a toy only to drop it) increases the likelihood they will develop a compulsion to continue to pursue the reward. *See generally* NATASHA DOW SCHÜLL, ADDICTION BY DESIGN: MACHINE GAMBLING IN LAS VEGAS (2014) (explaining presence of "near misses" and other operant conditioning mechanisms in gambling); *cf.* Deeb v. Stoutamire, 53 So. 2d 873, 875 (Fla. 1951) ("It is our thought that the element of unpredictability is not supplied because a player may not be sure what score he can accomplish, but that it must be inherent in the machine. Parenthetically, if he could be sure, why would he play?").

[32] U.S. OFF. OF THE SURGEON GEN., SOCIAL MEDIA AND YOUTH MENTAL HEALTH: THE U.S. SURGEON GENERAL'S ADVISORY 6–12 (2023), https://www.hhs.gov/sites/default/files/sg-youth-mental-health-social-media-advisory.pdf.

parenting, driving, and medical procedures.[33] Meanwhile, scholars such as Gaia Bernstein and Nita Farahany have recently pointed out that emerging technologies, from AI to neural implants, open up new possibilities for addictive design to be even more effective and profitable in the future.[34] Congress, too, has held multiple hearings touching on this issue.[35]

Of course, the idea that addiction and compulsion create a market failure that threatens public health—that "sellers" can exploit "buyers" by getting them hooked on their products—is not new (would that it were).[36] The traditional American response to this public health challenge has been for we the people, through our state legislatures (and courts adjudicating state law claims) to create (or apply) consumer protections for addictive products. Thus, states regulated alcohol,[37] tobacco,[38] and gambling[39] long before the federal government, and they

---

[33] *E.g.*, Erez Kita & Gil Luria, *The Mediating Role of Smartphone Addiction on the Relationship Between Personality and Young Drivers' Use While Driving*, 59 TRANSP. RSCH. 203, 203 (2018); Denise Ante-Contreras, Distracted Parenting: How Social Media Affects Parent-Child Attachment (2016) (M.S.W. thesis, California State University, San Bernadino), https://scholarworks.lib.csusb.edu/etd/292; Huseyin Ulas Pinar, Omer Karaca, Rafi Dogan & Ummu Mine Konuk, *Smartphone Use Habits of Anesthesia Providers During Anesthetized Patient Care: A Survey from Turkey*, 16 BMC ANESTHESIOLOGY 88, 89–91 (2016) ("93.7% of respondents used smartphones during anesthetized patient care" in operating rooms; 41% reported having seen such use by a colleague negatively impact care).

[34] *See generally* BERNSTEIN, *supra* note 4; FARAHANY, *supra* note 30.

[35] *E.g.*, *Protecting Kids Online: Testimony from a Facebook Whistleblower*: *Hearing Before the S. Comm. on Com., Sci. & Transp.*, 117th Cong. (2021), https://www.commerce.senate.gov/2021/10/protecting%20kids%20online:%20testimony%20from%20a%20facebook%20whistleblower.

[36] *See* Daniel J. Hemel & Lisa Larrimore Ouellette, *Innovation Institutions and the Opioid Crisis*, 7 J.L. & BIOSCIENCES 1 (2020) ("obviously, OxyContin generates negative externalities"). For an early work that influenced economists' understanding of addiction, see Gary S. Becker & Kevin M. Murphy, *A Theory of Rational Addiction*, 96 J. POL. ECON. 675, 691 (1988) ("Temporary events can permanently 'hook' rational persons to addictive goods.").

[37] State interest in regulating alcohol dates back to Founding Father Benjamin Rush's influential tract framing alcohol addiction as a physical rather than a moral problem, developed through exposure rather than vice. *See* Paul Aaron & David Musto, *Temperance and Prohibition in America: A Historical Overview*, *in* ALCOHOL AND PUBLIC POLICY: BEYOND THE SHADOW OF PROHIBITION 127, 149 (Mark H. Moore & Dean R. Gerstein eds., 1981) (describing Rush's tract); *see generally* Philip L. Hersch & Jeffry M. Netter, *State Prohibition of Alcohol: An Application of Diffusion Analysis to Regulation*, 12 RSCH. L. & ECON. 55 (1989) (describing adoption of state alcohol prohibition prior to federal prohibition).

continue to tinker with these protections to this day. Indeed, the Twenty-First Amendment (which repealed federal prohibition) is a rare constitutional provision specifically entrenching state authority over a particular subject matter—the regulation of alcohol—forbidding the importation of "intoxicating liquors" in violation of the laws of a state.

This longstanding tradition of state public health regulation of addictive products has seen states experiment (and learn from each other's experimentation) with a range of regulatory tools to address a vexing problem. This includes bans on the sale of particular products in general or to particularly vulnerable users (especially children and adolescents).[40] It also includes liability for steps taken by manufacturers to increase the addictiveness of their products, such as manipulating the content of cigarettes to increase their addictiveness.[41] Such civil litigation for violation of state common law and statutory requirements has proven a potent tool for both uncovering the extent of addictive design (since evidence may

---

[38] Rosalie Liccardo Pacula et al., *Developing Public Health Regulations for Marijuana: Lessons from Alcohol and Tobacco*, 104 AM. J. PUB. HEALTH 1021, 1021 (2014) (discussing examples of historical regulation).

[39] At the time of the ratification of the Fourteenth Amendment, two-thirds of states included prohibitions on lottery gambling in their constitutions. *See* Steven G. Calabresi & Sarah E. Agudo, *Individual Rights Under State Constitutions When the Fourteenth Amendment Was Ratified in 1868: What Rights Are Deeply Rooted in American History and Tradition?*, 87 TEX. L. REV. 7, 101 (2008). Such provisions were enacted not only due to corruption concerns but also due to concerns about the effect of gambling on the "character" of the population. *E.g.*, J. ROSS BROWNE, REPORT OF THE DEBATES OF THE CONVENTION OF CALIFORNIA, ON THE FORMATION OF THE STATE CONSTITUTION, IN SEPTEMBER AND OCTOBER, 1849, at 91 (1850) ("[Gambling] penetrates to the domestic circle . . . destroy[ing] the happiness of families, and fall[ing] with a particular weight upon the widow and the orphan.") (statement of Rep. Hoppe). On the role of federalism in gambling regulation more generally, see Murphy v. Nat'l Collegiate Athletic Ass'n, 138 S. Ct. 1461 (2018) (discussing role of federalism in gambling regulation); Nelson Rose, *Gambling and the Law: The Third Wave of Legal Gambling*, 17 VILL. SPORTS & ENT. L.J. 361, 365 (2010) (general overview of the history of gambling in American law).

[40] *E.g.*, Michael A. Wagner, *'As Gold Is Tried in the Fire, So Hearts Must Be Tried By Pain': The Temperance Movement in Georgia and the Local Option Law of 1885*, 93 GA. HIST. Q. 30 (2009) (discussing enactment in Georgia of law permitting alcohol prohibition at the local level in 1880).

[41] United States v. Phillip Morris USA, Inc., 449 F. Supp. 2d 1, 39 (D.D.C. 2006) (discussing tobacco companies' use of design features and chemical additives in the manufacturing process to more effectively "create and sustain addiction").

be closely held by manufacturers unless revealed through discovery) and for safeguarding public health.[42] Finally, state regulation of addictive products includes taxes on addictive products used to fund addiction awareness and treatment programs.[43] Indeed, a significant body of research associated with "legal epidemiology" uses experience with historical state regulation to help states refine and calibrate their regulations going forward.[44]

Given public health concerns surrounding addictive design, the history of state public health regulation of addictive products, and the Brandeisian benefits of state experimentation as a precursor to any uniform federal policy,[45] it is not surprising that states have been at the forefront of thinking about whether and how legal tools should come into play to safeguard the public's health against the alleged harms of addictive design by platforms and other digital technologies.[46] A growing list of states including California, Maryland, Minnesota, and New Jersey have considered or enacted online consumer protection bills that, to some extent, reflect public health concerns surrounding addictive design.[47]

To be sure, concerns about addictive design online are not without their skeptics who deny that addictive design is a genuine public health threat or, at least, deny that there is sufficient "gold standard," randomized double-blind trial evidence establishing such a threat (or that state intervention carries sufficient benefits) to justify regulatory intervention in the market. The same was true, for a time,

---

[42] Nora Freeman Engstrom & Robert L. Rabin, *Pursuing Public Health Through Litigation*, 73 STAN. L. REV. 285 (2021).

[43] *See generally* Andrew J. Haile, *Sin Taxes: When the State Becomes the Sinner*, 82 TEMP. L. REV. 1041, 1044 (2009) ("Taxes on harmful products have existed almost since the country's founding, and the debate over the virtues and vices of sin taxes are just as old.").

[44] *E.g.*, Scott Burris, Lindsay K. Cloud & Matthew Penn, *The Growing Field of Legal Epidemiology*, 26 J. PUB. HEALTH MGMT. & PRAC. S4 (2020).

[45] *See* New State Ice Co. v. Liebmann, 285 U.S. 262, 311 (1932) (Brandeis, J.) ("It is one of the happy incidents of the federal system a single courageous State may, if its citizens choose, serve as a laboratory; and try novel social and economic experiments without risk to the rest of the country."). On federal laws that block state experimentation without creating a federal regulatory regime (and thereby stifle not only state experimentation but also political pressure on the federal government to adopt a uniform regulatory approach, potentially entrenching a regulatory vacuum), see Jonathan Remy Nash, *Null Preemption*, 85 NOTRE DAME L. REV. 1015 (2010).

[46] Langvardt, *supra* note 4; Morgan, *supra* note 4.

[47] DiFilippo, *supra* note 2.

of public health concerns about the addictiveness of cigarettes and prescription opioids.[48] In these contexts, courts adjudicating tobacco and opioid claims played an important role as a forum for collecting and assessing the evidence to adjudicate the legitimacy of disputed claims.[49]

When it comes to disputes about the legitimacy of concerns about addictive design by technology companies, one might assume that courts will play the same information-forcing and adjudicatory role they have traditionally played with regard to public health threats in the United States. But such an assumption would be too hasty. Addictive design online is not only a public health regulatory question; it is also an internet regulatory question. And when it comes to the internet, courts' role has largely been to protect innovation by insulating platforms and providers from defending liability claims altogether, regardless of the merits.

B. Section 230 and the First Amendment May Limit State Authority

Faced with threatened state regulation, platforms have argued—and here (as throughout this Article) I articulate my understanding of their legal position and its import[50]—that states lack authority to regulate addictive design by providers of interactive computer services at all, no matter whether they actually know or intend to develop addictive products, no matter how severe any foreseeable harms may be, and no matter how easy it would be to tweak a product to reduce or eliminate any addictive potential. Addictive design online is a public-health-law-free zone, as I understand the platforms' view, because allegedly addictive design online necessarily implicates expression—both the expression of platforms (or other providers) and the expression (often called "content") of their users. Thus, the platforms argue, state authority in this space is preempted by both Section 230 and the First Amendment.[51]

Whether or not their expressive aspects actually make addictive platforms and games different in kind from other addictive products (especially slot machines

---

[48] Engstrom & Rabin, *supra* note 42, at 304, 307.

[49] *Id.*

[50] For more on my understanding of the platforms' position, see *infra* Parts II and III and accompanying text.

[51] *See infra* Part II.

and other gambling products) is debatable.[52] Suffice it to say, however, that there is no guarantee that courts will see state regulation of addictive design by platforms as a "public health" question—akin to state regulation of tobacco or gambling—as to which broad state public health power to regulate is the norm and federal limitations are viewed with some skepticism.[53] Quite the contrary, courts might instead see state regulation of addictive design by platforms online as an "internet regulation" question—akin to state regulation of content moderation choices whether and when to censor "fake news" and the like—as to which broad federal preemption (both statutory and constitutional) of state authority is the norm.[54]

States devising new laws and precedents, internet companies devising new technologies and policies, and we the people (aka "users") deciding how to protect our and our children's mental health (including freedom of thought) thus face an open legal question with significant implications for the future of an industry, for federalism, and for Americans' public health: When the dust settles, what authority will—or should—states have to regulate addictive design online?

Litigation on these questions is no longer merely hypothetical. The first major cases to test them broadly, *In re Social Media Adolescent Addiction Litigation* and

---

[52] Sale and use of more familiar addictive products like cocaine or tobacco can implicate expression (among other constitutionally protected interests), *see* 1 MARC JONATHAN BLITZ & JAN CHRISTOPH BUBLITZ, THE LAW AND ETHICS OF FREEDOM OF THOUGHT 15–16 (2021) (describing argument and collecting sources); Jan Christoph Bublitz, *My Mind is Mine!? Cognitive Liberty as a Legal Concept*, in COGNITIVE ENHANCEMENT (Elisabeth Hildt & Andreas Francke eds., 2013). Gambling has long been regulated despite the fact that, from the flashing lights and rolling wheel of a slot machine to the competitive camaraderie of a poker table, gambling can be said to be inextricably expressive. That the question whether the promotion of gambling itself gets the benefit of First Amendment protection has been controversial, *see* United States v. Edge Broad. Co., 509 U.S. 418, 436 (1993) ("Because the statutes challenged here regulate commercial speech in a manner that does not violate the First Amendment, the judgment of the Court of Appeals is Reversed."), merely illustrates the longstanding acceptance of the proposition that state regulation of gambling itself is not subject to ordinary First Amendment strictures. *Cf.* Evans, *supra* note 21, at 414 (noting that case law supports recognition of loot boxes as a form of gambling); *id.* at 420 (discussing the social costs of loot boxes *vis-à-vis* their nature as a type of gambling).

[53] *See* Jacobson v. Massachusetts, 197 U.S. 11, 13 (1905); Oregon v. Ashcroft, 368 F.3d 1118, 1124 (9th Cir. 2004), *aff'd sub nom.* Gonzales v. Oregon, 546 U.S. 243 (2006).

[54] 47 U.S.C. § 230(e)(3) ("No cause of action may be brought and no liability may be imposed under any State or local law that is inconsistent with this section.").

the *Social Media Cases*,[55] offer a useful opportunity to begin to map this next frontier of public health law.

C. Issue Is Joined in In re Social Media Adolescent Addiction Litigation

In re Social Media Adolescent Addiction Litigation (or *In re Social Media Addiction* for short)[56] is a consolidated case grouping numerous lawsuits alleging that major platforms (Facebook, Instagram, Snapchat, TikTok, and YouTube) designed their products to foster compulsion in adolescents—and failed to warn about this—which promoted user numbers and "time on platform" while harming plaintiffs in the process.[57] The alleged harms suffered by the various plaintiffs are analogous to the harms described in the 2023 Surgeon General's report warning parents about social media use by their children.[58] They are also consistent with testimony in congressional hearings on the interaction between social media and mental health.[59]

At its core, the complaint in *In re Social Media Addiction* alleges that the platforms "wrote code designed to manipulate dopamine release in children's devel-

---

[55] *See* Sharyn Alfonsi, *More Than 1,200 Families Suing Social Media Companies over Kids' Mental Health*, CBS NEWS (Dec 11, 2022, 3:58 PM), https://www.cbsnews.com/sacramento/news/social-media-lawsuit-meta-tiktok-facebook-instagram-60-minutes-2022-12-11/?intcid=CNM-00-10abd1h ("Today, there are more than 1,200 families pursuing lawsuits against social media companies including TikTok, Snapchat, YouTube, Roblox and Meta, the parent company to Instagram and Facebook. More than 150 lawsuits will be moving forward next year.").

[56] *See In re* Soc. Media Adolescent Addiction/Pers. Inj. Prods. Liab. Litig., No. 4:22-MD-03047-YGR, 2023 WL 2414002, at *1 (N.D. Cal. Mar. 8, 2023).

[57] For failure to warn allegations, see Compl., *supra* note 18, at ¶ 238 ("Facebook and Instagram owe their success to their defective design, including their underlying computer code and algorithms, and to Meta's failure to warn plaintiffs and Consortium plaintiffs that the products present serious safety risks. Meta's tortious conduct begins before a user has viewed, let alone posted, a single scrap of content."); *see also id.* ¶¶ 434, 816, 909. For allegations of failure to warn based on a theory of strict liability, see *id.* at 242. For negligence-based claims, see *id.* at 248. For claims asserting that defendant platforms' products are "designed to addict," see *id.* ¶¶ 467, 840, 864.

[58] U.S. OFF. OF THE SURGEON GEN., *supra* note 32, at 6–12.

[59] *See Protecting Kids Online*, *supra* note 35; *cf. Tobacco CEO's Statement to Congress 1994 New Clip "Nicotine Is Not Addictive,"* UCSF ACAD. SENATE, https://senate.ucsf.edu/tobacco-ceo-statement-to-congress.

oping brains and, in doing so, create compulsive use of their apps."[60] As the plaintiffs put it:

> Borrowing heavily from the behavioral and neurobiological techniques used by slot machines and exploited by the cigarette industry, Defendants deliberately embedded in their products an array of design features aimed at maximizing youth engagement to drive advertising revenue. Defendants know children are in a developmental stage that leaves them particularly vulnerable to the addictive effects of these features. Defendants target them anyway.[61]

They allege that the platforms are addictive as a whole and that particular features of their design contribute individually and collectively to this addictiveness. Plaintiffs allege that the platforms not only use the same operant conditioning tricks that can make slot machines addictive,[62] they improve upon them:

> [S]lot machines [are] limited by the fact that they deliver rewards . . . irrespective of the person pulling the lever. By contrast, Defendants' apps are designed to purposely withhold and release rewards on a schedule its algorithms have determined is optimal to heighten a specific user's craving and keep them using the product. For example . . . Instagram's notification algorithm will at times determine that a particular user's engagement will be maximized if the app *withholds 'Likes' on their posts* and then later delivers them in a large burst of notifications.[63]

Additionally, plaintiffs allege "dangerous and defective features," including pushing users toward particularly stimulating content,[64] building structures that use social feedback to prompt and retain engagement,[65] encouraging the creation and sharing of content that entails unrealistic, "filtered and fake appearances and experiences,"[66] failing to provide effective parental controls (such as minimal age

---

[60] Compl., *supra* note 18, ¶ 12.

[61] *Id.* ¶ 2.

[62] *Id.* ¶ 79.

[63] *Id.* ¶ 79.

[64] *Id.* ¶ 82.

[65] *Id.* ¶ 86 ("For example, in the real world, no public ledger tallies the number of consecutive days friends speak. Similarly, 'after you walk away from a regular conversation, you don't know if the other person liked it, or if anyone else liked it.' By contrast, a product defect like the 'Snap Streak' creates exactly such artificial forms of feedback." (quoting Zara Abrams, *Why Young Brains Are Especially Vulnerable to Social Media*, Am. Psych. Ass'n (Aug. 25, 2022))).

[66] *Id.* ¶ 88.

verification),[67] making it extremely difficult to "quit" the platforms,[68] encouraging dangerous "challenges" (such as the "Blackout challenge" that became popular on TikTok),[69] and contributing to the sexual exploitation of children.[70]

As early, pre-discovery support of their allegations, the plaintiffs in *In re Social Media Addiction* offer citations to neuroscience and psychology literature, citations to federal reports and hearings from the Surgeon General and Congress,[71] and a litany of quotes from internal documents developed by some platforms, as well as former platform officials. For example, plaintiffs allege that an internal Facebook study of "problematic users" noted that "[a]ll problematic users were experiencing multiple life impacts" including "loss of productivity, sleep disruption, relationship impacts, and safety risks."[72] And Facebook's first President, Sean Parker, said the following in an interview in 2017:

> God only knows what it's doing to our children's brains. . . . The thought process that went into building these applications, Facebook being the first of them, . . . was all about: "How do we consume as much of your time and conscious attention as possible?" . . . And that means that we need to sort of give you a little dopamine hit every once in a while, because someone liked or commented on a photo or a post. . . . And that's going to get you to contribute more content, and that's going to get you . . . more likes and comments . . . It's a social-validation feedback loop . . . exactly the kind of thing that a hacker like myself would come up with, because you're exploiting a vulnerability in human psychology. . . . The inventors, creators—it's me, it's Mark [Zuckerberg], it's Keven Systrom on Instagram, it's all of these people—understood this consciously. And we did it anyway.[73]

Pointing out spiking rates of youth suicide and mental illness and polls showing huge numbers of adolescents who report using social media "too much" or

---

[67] *E.g., id.* ¶¶ 328–29.

[68] *E.g., id.* ¶ 358.

[69] *Id.* ¶ 126.

[70] *Id.* ¶ 133–35.

[71] U.S. Surgeon General, *supra* note 58, at 6–12.

[72] Compl., *supra* note 18, ¶ 376 n.84.

[73] *Id.* ¶ 261 (quoting Mike Allen, *Sean Parker Unloads on Facebook: "God Only Knows What It's Doing to Our Children's Brains,"* AXIOS (Nov. 9, 2017), https://www.axios.com/2017/12/15/sean-parker-unloads-on-facebook-god-only-knows-what-its-doing-to-our-childrens-brains-1513306792).

find it hard to quit, the plaintiffs allege the platforms have created "nothing short of a national crisis."[74]

As for claims, the plaintiffs press numerous distinct counts. These include strict liability and negligence claims for design defects, strict liability and negligence claims for failure to warn, violation of unfair trade practices and consumer protection laws, fraudulent and negligent concealment and misrepresentation, negligence per se, wrongful death, survival, loss of consortium, and violations of federal statutes related to child trafficking.[75] For example, plaintiffs' strict liability design defect claim alleges that "[e]ach of the Defendant's [sic] defectively designed its respective products to addict minors and young adults, who were particularly unable to appreciate the risks posed by the products, and particularly susceptible to harms from those products,"[76] and that "each Defendant knew, or by the exercise of reasonable care, should have known" about the same.[77] Their failure to warn claim alleges, *inter alia*, that the platforms "failed to exercise reasonable care to inform users that . . . [their] products cause addiction, compulsive use, and/or other concomitant physical and mental injuries."[78]

The defendants in *In re Social Media Addiction* moved to dismiss on various merits grounds particular to plaintiffs' specific claims. Separate from that, they also moved to dismiss all claims—even those as to which their motion to dismiss on the particulars fails—on two global grounds.[79] First, they assert that all claims are federally preempted in their entirety by Section 230.[80] Second, they assert that, as challenges to ultimately expressive activity, all claims (or rather, the state common law and consumer protection laws underlying the claims, as applied) violate

---

[74] *Id.* ¶¶ 117, 301.

[75] *See* 18 U.S.C. §§ 1595, 2255, 2252A, 2255, 2258B.

[76] Compl., *supra* note 18, ¶ 837.

[77] *Id.* ¶ 839.

[78] *Id.* ¶ 864.

[79] Defendants' Supplemental Joint Motion to Dismiss Based on Section 230 and the First Amendment at 2, *In re* Soc. Media Adolescent Addiction/Pers. Inj. Prods. Liab. Litig., No. 4:22-MD-03047-YGR (N.D. Cal. June 27, 2023) [hereinafter Defs.' Mot.] (Section 230-based arguments for blanket dismissal); *id.* at 14 (First Amendment-based arguments for blanket dismissal).

[80] *Id.* at 2.

the First Amendment.[81] At this writing, the district court in *In re Social Media Addiction* has denied in part and granted in part the platforms' motion.

The court's resolution of the motion to some extent tracks the October 2023 resolution of a related motion in another addictive design proceeding, the *Social Media Cases*, in California state court. The court there issued a first opinion on the reach of Section 230 and the First Amendment as to addictive design claims. These early cases offer an early, useful, and timely opportunity to analyze state authority to regulate addictive design online. Because the platforms' Section 230 and constitutional objections largely attack the extent of state authority in this space, not just the specifics of plaintiffs' claims, the ultimate resolution of those issues may well determine where, when, and how states may regulate addictive design more generally—whether through common law torts, generally applicable consumer protection statutes, or specific legislation directed at public health concerns related to addictive design.

II. CHARTING SECTION 230

To help readers form a picture of the paths courts might ultimately take in determining state authority to regulate addictive design, this part begins by describing and generalizing the parties' respective theories of the scope of Section 230 in *In re Social Media Adolescent Addiction Litigation* as well as Judge Kuhl's initial resolution of preemption arguments in the *Social Media Cases*.[82] Subpart A offers background on Section 230 and discusses the platforms' theory, on which state authority to regulate addictive design hinges on the larger "matchmaking" question of the extent of state authority to regulate content moderation considered but not resolved by the Supreme Court in *Gonzalez*. Subpart B describes the plaintiffs' theory of the limits of Section 230, which would permit broad state authority to regulate "design" claims. Subpart C describes and defends the third theory of the limits of Section 230 adopted by Judge Kuhl—that the statute is inapplicable to laws regulating platforms' content-neutral activities. This theory leaves states at least some authority to regulate addictive design regardless of how they resolve the *Gonzalez*/matchmaking issue.

---

[81] *Id.* at 14.

[82] I don't separately analyze the arguments in the briefs in the *Social Media Cases* because they largely overlap.

A. Moderation vs. Matchmaking

1. The *Gonzalez*/matchmaking debate

Because Section 230 is a pivotal limit preempting state authority online (and the centerpiece of the platforms' motion to dismiss), some brief background is warranted. There is consensus that Section 230(c)(2) protects platforms (and other providers) from liability for affirmatively censoring particular content (though there is some debate about the extent of that protection).[83] There is also consensus that Section 230(c)(1) protects platforms from liability for *failing* to censor content.[84] Harder questions emerge when a platform's conduct entails not absolutely foreclosing access to content (or failing to do so), but instead influencing the likelihood that a particular user (or users) will view particular content by making prioritization choices. Prioritization choices might include choices about what content to feature at the "top" of the site's landing page (equivalent to a newspaper's decision about what headline to place "above the fold") or what content to recommend a user "might like" (equivalent to one friend suggesting a book based on her intimate knowledge of another friend's tastes).

There is at this writing a significant, unresolved legal controversy about the extent to which Section 230(c)(1) bars regulation of website conduct to prioritize, recommend, or otherwise steer users toward particular content.[85] Conflicting views on this question have emerged in the circuits and were debated before the Supreme Court in *Gonzalez v. Google*—though the high court did not ultimately rule on the question and oral argument suggested that the conceptual issues it raised need further development and clarification, especially to determine what line might separate unavoidable prioritization protected by Section 230 from more affirmative recommendation or matchmaking that might fall outside Section 230's protection.[86]

---

[83] *See* Adam Candeub & Eugene Volokh, *Interpreting 47 U.S.C. § 230(c)(2)*, 1 J. FREE SPEECH L. 175 (2021).

[84] *E.g.*, Brief for the United States as Amicus Curiae at 24, Gonzalez v. Google LLC, 598 U.S. 617 (2023) ("Section 230(c)(1) bars plaintiffs' claims to the extent they are premised on YouTube's failure to block or remove third party content.").

[85] *E.g.*, *id.* at 26 ("Section 230(c)(1) does not preclude plaintiffs' claims based on YouTube's targeted recommendations.").

[86] Transcript of Oral Argument at 45, Gonzalez v. Google LLC, 598 U.S. 617 (2023) (Kagan, J.) ("And, you know, every other industry has to internalize the costs of its conduct. Why is it that

A view most prominently associated with an influential concurrence by the late great Judge Katzmann holds that at some point, a website is no longer merely displaying content submitted by third parties but actively playing the role of "matchmaker," connecting users to particular content based on the likelihood it will interest them (by considering all manner of individualized data about them and the content with which they are matched), and that Section 230 protection ceases at this point.[87] A competing view holds that platforms (like newspapers) have no choice but to engage in some decisionmaking about what content to prioritize for users, that the protection of Section 230 would be eviscerated if claims for "failing to censor" could be re-conceptualized as claims for "prioritizing," and that no meaningful distinction may be drawn between the sort of unavoidable prioritization choices platforms must make and more "targeted" or affirmative "recommendations" that would permit the latter to escape the protection of Section 230 without exposing the former.[88] The Ninth Circuit largely adopted the latter view in *Gonzalez v. Google*, holding that YouTube's recommendation of ISIS videos to individuals who ultimately (and allegedly) became radicalized to commit acts of terrorism was protected by Section 230(c)(1),[89] though that precedent was vacated as a result of the Supreme Court's resolution of the case on other grounds.[90]

---

the tech industry gets a pass? A little bit unclear. On the other hand, I mean, we're a court. We really don't know about these things. You know, these are not like the nine greatest experts on the Internet. And I don't have to—I don't have to accept all Ms. Blatt's 'the sky is falling' stuff to accept something about, boy, there is a lot of uncertainty about going the way you would have us go, in part, just because of the difficulty of drawing lines in this area and just because of the fact that, once we go with you, all of a sudden we're finding that Google isn't protected. And maybe Congress should want that system, but isn't that something for Congress to do, not the Court?").

[87] Force v. Facebook, Inc., 934 F.3d 53 (2d Cir. 2019).

[88] Brief for Respondent at 33, 2022 WL 18358194, Gonzalez v. Google LLC, 598 U.S. 617 (2023) ("Every claim could be recast as challenging how websites sort and prioritize third-party content. TripAdvisor might be sued for tortious interference with business relations by prominently listing one-star reviews. Lexis might be sued for contributing to defamation by prioritizing a defamatory law-review article. . . . Given that virtually everyone depends on tailored online results, Section 230 is the Atlas propping up the modern internet—just as Congress envisioned in 1996.").

[89] Gonzalez v. Google LLC, 598 U.S. 617, 621 (2023).

[90] *See id.*

2. Challenges to understanding addictive design within the matchmaking debate

The platforms frame the claims in *In re Social Media Addiction* within this broader debate about the extent to which Section 230 applies to content recommendations. They argue the case is governed by Ninth Circuit precedent predating *Gonzalez* (especially *Dyroff*),[91] in which precedent seemingly resolves the matchmaking issue in favor of preemption.[92] To some extent, they have a real point: Allegations that the platforms' products caused compulsion or other public health harms by, *e.g.*, recommending stimulating or aversive content,[93] are difficult to distinguish from allegations in *Gonzalez* that YouTube caused harm by recommending ISIS videos or allegations in *Dyroff* that the Experience Project website caused harm by connecting an adolescent with a drug dealer.

The problem with the platforms' effort to frame addictive design claims within *Dyroff* (and the broader dispute about the applicability of Section 230 to content moderation) is that while some allegations relating to addictive design indeed depend on the particular content recommended by platforms, most do not. The *Gonzalez*/matchmaking issue, while fascinating and certainly relevant to the *In re Social Media Addiction* case, is therefore in some sense a red herring. Many of the plaintiffs' claims do not seem to involve content prioritization or recommendation at all. As discussed further in Part IV, these include plaintiffs' claims that the platforms built intermittent reinforcement and variable reward features into their products, that they built structures that use social feedback to prompt and retain engagement,[94] that they failed to provide a warning or effective parental controls

---

[91] *Dyroff* found that Section 230 foreclosed liability for recommendations that matched an adolescent with a fentanyl dealer, and that ultimately led to the adolescent's death from a drug overdose. Dyroff v. Ultimate Software Grp., Inc., 934 F.3d 1093, 1096 (9th Cir. 2019).

[92] *E.g.*, Defendants' Reply in Support of Supplemental Joint Motion to Dismiss Pursuant to Rule 12(b)(6) Plaintiffs' Priority Claims Under Section 230 and the First Amendment at 3–4, *In re* Soc. Media Adolescent Addiction/Pers. Inj. Prods. Liab. Litig., No. 4:22-MD-03047-YGR-TSH (N.D. Cal. Aug. 15, 2023) [hereinafter Platforms' Reply] (rulings related to Section 230's application to content recommendations that "wholly refute" plaintiffs' claims).

[93] Compl., *supra* note 18, ¶ 82.

[94] *Id.* ¶ 86 ("For example, in the real world, no public ledger tallies the number of consecutive days friends speak. Similarly, 'after you walk away from a regular conversation, you don't know if the other person liked it, or if anyone else liked it.' By contrast, a product defect like the 'Snap Streak' creates exactly such artificial forms of feedback." (quoting Zara Abrams, *Why Young*

(such as minimal age verification),[95] and that they have made it needlessly difficult to "quit."[96]

The platforms' brief obscures this challenge with wordplay. The platforms assert that Section 230 applies not only to claims that "focus on a particular piece of content," but also to claims that "referenc[e] general categories of potentially harmful content,"[97] and still further, to all claims for which content is an "indispensable"[98] or "inextricabl[e]"[99] part, even if such claims relate to "all user content available" on a service rather than "just some" content.[100] In other words, and as I understand it, the platforms in *In re Social Media Addiction* argue that Section 230 applies to all claims that relate to content (that is, to which content is "indispensable" or "inextricable")— not just claims that relate to *particular* posts or *particular* categories of content, but to claims that relate to all material that is or might be shared on a platform, regardless of its content.

So understood, there are several serious obstacles to the platforms' theory of the case. First, in their reply brief, the platforms do not draw the key, operative terms that extend Section 230 beyond core content moderation activities—that is, their assertions that Section 230 applies if content is an "inextricable" or "indispensable" part of the basis for liability—from caselaw. The closest they come to citing precedent for this leap is the Fifth Circuit's rejection of what it saw as an effort to plead around Section 230 by framing a claim for connecting an adolescent with a sexual predator as a claim for products liability (for having built a machine that could connect an adolescent with a sexual predator) in *Doe v. MySpace, Inc.*[101] As the platforms point out, the Fifth Circuit rejected that effort as "merely

---

Brains Are Especially Vulnerable to Social Media, AM. PSYCH. ASS'N (Aug. 25, 2022), https://www.apa.org/news/apa/2022/social-media-children-teens)).

[95] *E.g.*, Compl., *supra* note 18, ¶¶ 328–29.

[96] *E.g.*, *id.* ¶ 358.

[97] Platforms' Reply, *supra* note 92, at 3.

[98] *Id.* at 2.

[99] *Id.* at 7.

[100] Defendants' Joint Motion to Dismiss Pursuant to Rules 12(b)(1) and 12(b)(6) Plaintiffs' Priority Claims Asserted in Amended Master Complaint at 26, *In re* Soc. Media Adolescent Addiction/Pers. Inj. Prods. Liab. Litig., No. 4:22-MD-03047-YGR (N.D. Cal. Apr. 17, 2023); Platforms' Reply, *supra* note 92, at 3.

[101] 528 F.3d 413, 420–22 (5th Cir. 2008).

another way of claiming that MySpace was liable for publishing the communications."[102] But the point that a challenge to a platform's failure to discriminate among content may be barred by Section 230 whether framed as a direct attack on the sharing of particular harmful content or a more abstract, products liability claim for having built a product that could share such content, tells us nothing about whether Section 230 applies to platforms' *content-neutral* decisions regulating the time, place, and manner of user speech. The platforms did not draw the terms "inextricable" and "indispensable" from the *Doe* opinion, or any other source in Section 230 case law.[103]

The lack of precedent for the platforms' assertion that Section 230 blocks any claim to which content is "indispensable" or "inextricable" is particularly notable because of the extent to which, if accepted, it would insert the hands-off internet paradigm in place of the traditional, Brandeisian (that is, led by state innovation) public health paradigm by broadly preempting state authority to regulate addictive design. It is difficult to see what space, if any, states would have to regulate addictive design by entities subject to Section 230's protections if courts were to adopt this argument—no matter the potency of current conditioning techniques or those that might be developed in the future, or the strength of the public health evidence base supporting concerns about their mental health impacts.

Though the platforms do not say so explicitly, the clear implication of their arguments is that even if (as alleged) they really did knowingly (or at least negligently) design their products to be addictive,[104] and even if (as alleged) kids in a state suffered severe harm as the result of the compulsions they unwittingly developed to use the platforms' products,[105] and even if (as alleged) such conduct

---

[102] Platforms' Reply, *supra* note 92, at 7 (quoting *Doe*, 528 F.3d at 420).

[103] *Id.* at 2, 7. A Westlaw search of all federal cases for "*section 230*" *&* *((inextricable /4 content) (indispensable /4 content))* yields zero results as of October 6, 2023. Now that Judge Koh has ruled on the platforms' motion to dismiss, the term "inexplicable" does appear in section 230 case law in the context of her description and rejection of their theories. *In re* Soc. Media Adolescent Addiction/Pers. Inj. Prod. Liab. Litig., No. 4:22-MD-03047-YGR, 2023 WL 7524912, *16 (N.D. Cal. Nov. 14, 2023) ("Defendants' . . . assert[ion] that [the negligence *per se* claim] is barred because the harm alleged is inextricable from third-party content . . . is not an adequate basis for section 230 immunity.").

[104] *E.g.*, Compl., *supra* note 18, ¶ 12.

[105] *E.g.*, *id.* ¶ 117.

would violate state law (including both common law tort and consumer protection statutes), there would be nothing the state could do about it. Indeed, on the platforms' view, states could not even impose warning requirements[106] (akin to those found on tobacco products) or insist on effective parental controls (akin to limitations on the purchase of tobacco by juveniles) to make sure that users could make an informed choice about whether to expose themselves to an intentionally addictive technology, insofar as the consumption of content through the technology was an "indispensable" or "inextricable" part of any claim for violation of such laws.

Such a reading of Section 230 to broadly preempt states' traditional public health authority to regulate addictive products in an entire emerging field is arguably a "major question," such that courts should reject it absent a clear manifestation of congressional intent. In *Gonzalez v. Oregon*, the Supreme Court saw the tradition of state leadership in regulating access to potentially dangerous drugs as so well entrenched that it refused to read an alleged ambiguity in the Controlled Substances Act to empower the Attorney General to intrude on state authority in the absence of explicit textual support.[107] The Supreme Court has subsequently cited that case as an illustration of the canon that courts will not lightly read ambiguities in statutes to resolve "major questions."[108] This offers reason why courts should be especially reluctant to employ a reading of Section 230 that is not compelled by its text to displace traditional state authority to protect their residents from unwitting exposure to addictive products.

Such a reading of Section 230 to insulate addictive design would also undermine the purposes of Section 230. Addictive design is a threat not only to public health, but *also* a threat to innovation on the internet.[109] Today there is a risk that

---

[106] *See* Defs.' Mot., *supra* note 79, at 18 (arguing that "failure to warn" claims are barred by Section 230); *see also* Platforms' Reply, *supra* note 92, at 8–9.

[107] Gonzales v. Oregon, 546 U.S. 243, 274–75 (2006) ("The Government, in the end, maintains that the prescription requirement delegates to a single executive officer the power to effect a radical shift of authority from the States to the Federal Government to define general standards of medical practice in every locality. The text and structure of the CSA show that Congress did not have this far-reaching intent to alter the federal-state balance and the congressional role in maintaining it.").

[108] West Virginia v. EPA, 142 S. Ct. 2587, 2595 (2022).

[109] *See supra* note 29 and accompanying text (collecting sources).

the need for a high volume of users to harness network effects[110] gives platforms an economic incentive to compete not by offering better quality products, but by offering more addictive products (and getting the next generation of users hooked on them early).[111]

Given this concern about the distorting effect of addictive design on innovation online, state regulation of addictive design has the potential to further rather than undermine Section 230's purpose of "preserving a vibrant and competitive free market . . . for Internet and other interactive computer services."[112] In an unregulated environment, responsible platforms that seek to safeguard their users' mental health while offering innovative, value-adding features may quickly find themselves pushed out of the market by irresponsible platforms who develop a competitive advantage by tricking unwitting users into developing compulsions that bring volume, retention, and high time-on-device. By counteracting this dynamic, the regulation of addictive design has the potential to ensure a free market in which platforms compete to design the highest quality products, not the most addictive.[113]

Finally, the platforms' view of Section 230 is problematic not only due to its novelty and breadth. As the plaintiffs in *In re Social Media Addiction* point out, it also conflicts with existing precedent in the Ninth Circuit and other circuits.

B. Design vs. Derivative

In opposing the platforms' motion to dismiss in *In re Social Media Addiction*, the plaintiffs rely heavily on *Lemmon*, a Ninth Circuit case that they frame as precluding the platforms' theory that Section 230 applies to all content-related claims. In *Lemmon*, the Ninth Circuit held that Section 230 did not preclude a lawsuit alleging that Snapchat's "Speed Filter" had contributed to the death of several teenage boys. The boys used Snapchat and the "Speed Filter" to record themselves driving 120 miles per hour in a video shared through the app. While

---

[110] *See* TIM WU, THE CURSE OF BIGNESS: ANTITRUST IN THE NEW GILDED AGE (2018).

[111] *See supra* notes 28–29 and accompanying text.

[112] 47 U.S.C. § 230(b)(2).

[113] *See* 47 U.S.C. § 230(b)(4) (policy of the United States includes "remov[ing] disincentives for the development and utilization of blocking and filtering technologies that empower parents to restrict their children's access"); 47 U.S.C. § 230(a) (purpose to promote provision of "educational and informational resources").

filming, they veered off the road and suffered a fatal crash.[114] The plaintiffs alleged that the filter contributed to the accident, and the Ninth Circuit held that Section 230 did not bar the case.

Lemmon does seem to vitiate the argument that a claim's connection to content automatically brings it within the scope of Section 230. Content—the content submitted by the users—was inextricable (and indispensable) to the claims in the case. The "Speed Filter" didn't display *any* speed—it displayed the speed of the video content (any video content) uploaded by the plaintiff;[115] indeed, that was the problem. If the plaintiff hadn't been uploading their video content—if they had simply uploaded a blank box—there would not have been a case.[116]

Lemmon shows that some content-related claims are not barred by Section 230—but which ones? What divides content-related claims that are subject to Section 230 from content-related claims that fall within the *Lemmon* ambit and are not?

Plaintiffs offer the fact that the products liability claims in *Lemmon* were premised on Snapchat's own conduct (the design and offering of the "Speed Filter") as the key distinction. "Section 230 bars claims that seek to impose derivative liability for the content of third-party posts," they argue, "[i]t does not immunize conduct of the platforms themselves" (other than publishing or refusing to publish particular content).[117]

---

[114] Lemmon v. Snap, Inc., 995 F.3d 1085, 1088 (9th Cir. 2021) ("The app also permits its users to superimpose a 'filter' over the photos or videos that they capture through Snapchat at the moment they take that photo or video. Landen used one of these filters—the 'Speed Filter'—minutes before the fatal accident.").

[115] *Id.* ("The Speed Filter enables Snapchat users to 'record their real-life speed.'").

[116] The platforms attempt to distinguish *Lemmon* by pointing out that unlike the "Speed Filter" in that case, some features of their products challenged in *In re Social Media Addiction* involve a platform's recommendations about what third-party content users should view. Platforms' Reply, *supra* note 92, at 5–7. But as discussed above, other features challenged in the case do not involve such recommendations. The neutrality triangulation approach described *infra* Part II.C is a way to divide *Gonzalez/Dyroff*-type claims from *Lemmon*-type claims.

[117] Plaintiffs' Opposition to Defendants' Supplemental Motion to Dismiss Pursuant to Rule 12(b)(6) Plaintiffs' Priority Claims under Section 230 and the First Amendment at 3, In re: Social Media Adolescent Addiction/Personal Injury Products Liability Litigation, No. 4:22-MD-03047-YGR (N.D. Cal. July 25, 2023) [hereinafter Pls.' Opp.].

This theory—which largely limits Section 230 to what Evelyn Douek refers to as "first wave" (or direct) content moderation and leaves states empowered to regulate "second wave" (or systemic) content moderation[118]—seems plausible, but two challenges may undermine courts' willingness to accept this distinction. First, it may prove too much. As Douek suggests, any individual content moderation decision whether to censor, decline to censor, recommend, or decline to recommend a particular piece of content necessarily flows from a provider's upstream choices about how to administer content, *i.e.*, about the design of the product's content moderation apparatus.[119] As such, on this approach, it is hard to think of claims that cannot be cast as an upstream challenge to a product's design rather than as a downstream challenge to a particular content moderation choice (although doing so would of course require an underlying source of liability that was also focused on design rather than implementation).

Second, the distinction is incomplete, because *some* platform conduct is necessarily protected by Section 230, such as platform conduct that involves censoring posts or failing to censor posts. Indeed, in *Barnes* the Ninth Circuit held that Section 230 immunized Yahoo from a claim that it had negligently failed to remove certain explicit content[120]; such a claim necessarily targets the reasonableness of the platform's system for deciding when and how to remove content.

In implicit recognition of these challenges and addressing *Barnes*, plaintiffs offer the following to describe the upstream platform conduct that Section 230 does protect: "Section 230(c)(1) bars claims based upon duties that 'would necessarily require an internet company to monitor third-party content.'"[121] In other words, Section 230 does protect platform conduct in some cases, namely, the platform's conduct in failing to censor third-party content (or sharing it without monitoring it). It makes sense (and explains *Barnes*) that Section 230 protects such conduct, but it seems to me that framing this as the *outer bound* of platform conduct protected by Section 230 requires resolving the *Gonzalez* question.

---

[118] Evelyn Douek, *Content Moderation as Systems Thinking*, 136 HARV. L. REV. 526, 530 (2022) (advocating for a "systems thinking approach to content moderation regulation that focuses on systems rather than individual cases").

[119] *See generally id.*

[120] Barnes v. Yahoo!, Inc., 570 F.3d 1096, 1098 (9th Cir. 2009).

[121] Pls.' Opp., *supra* note 117, at 4 (quoting HomeAway.com, Inc. v. City of Santa Monica, 918 F.3d 676 (9th Cir. 2019)).

If courts hold (as the Solicitor General advocated in *Gonzalez*) that Section 230 protects only platform decisions whether to censor content or not—and does not protect platform decisions about content prioritization—then it would make sense to say, as do plaintiffs, that Section 230 applies to laws regulating platform conduct only if such laws would "necessarily require an internet company to monitor third-party content." But if, on the other hand, courts were to hold that Section 230 protects some content prioritization choices (some matchmaking or recommending), then it would necessarily reach laws regulating other aspects of platform conduct such as laws that would necessarily require a company to recommend (or decline to recommend) certain categories of content.

It might seem, therefore, that the ultimate resolution of the legal questions in *In re Social Media Addiction* (and so state authority to regulate addictive design) hinges on the *Gonzalez* question of whether Section 230 protects only content *removal* choices and not some content *prioritization* choices (or some design-focused version of it). Such a conclusion would be too hasty, however. There is a strong argument that some regulation of addictive design is consistent with Section 230 *even if* the *Gonzalez* question were to be resolved in favor of broader preemption and Section 230 read to protect platforms' content prioritization choices from state regulation.

C. Content Based vs. Content Neutral

Some platform (and other interactive computer service provider) conduct is protected by Section 230, and the question is what such conduct is protected. Plainly, content-independent conduct—conduct that has nothing at all to do with content—is not protected. And to the author it seems clear, too, that some content-related conduct *is not* protected—*Lemmon* (in the Ninth Circuit, at least) establishes this (among other considerations discussed below). But what is the line between content-related platform conduct that is protected by Section 230 and so beyond the reach of states, on the one hand, and content-related platform conduct that is not protected by Section 230 and so may be regulated by states, on the other?

Courts could conceivably draw these lines either by holding Section 230 does not reach "matchmaking" type decisions by platforms or by holding Section 230 does not reach "design" challenges. While I have noted limits to these approaches above, I don't mean to take a firm position on them here. I do wish to highlight, however, that even if courts reject both the "matchmaking" and "design" theories

of the limits of Section 230, significant aspects of addictive design would still fall outside of Section 230 insulation, because of another limit on the reach of Section 230 best understood by analogy to the First Amendment. Judge Kuhl's opinion in the *Social Media Cases* illustrates this approach.

At bottom, what we call "content" in Section 230 speak is essentially what constitutional law has long called "expression." And constitutional law has long separated often-permissible regulation of expression from usually-impermissible regulation of expression by looking to whether a regulation of expression is "content-based" (that is, it treats expression differently based on its content) or "content-neutral" (that is, it treats all expression the same even while regulating its time, manner, or place). Generally speaking, the law is much more comfortable with content-neutral state regulation of expression than it is with content-based state regulation of expression, because the latter has much less potential to let the state influence ideas and viewpoints.

In her October 2023 ruling in the *Social Media Cases*, Judge Kuhl followed the path of important prior cases in articulating a similar distinction in the Section 230 context, finding that Section 230 does not bar state regulation of content-neutral (though content-related) platform conduct.[122] This test makes sense and provides clear guidance about a core domain in which states may regulate addictive design notwithstanding ongoing uncertainty about Section 230's applicability to "matchmaking" and "design" conduct.

Section 1 below will use Judge Koh's decision in *In re Zoom Video Communications, Inc. Privacy Litigation* to introduce this approach because the opinion offers the fullest explanation of how it follows Section 230's text, purpose, history, and precedent. Section 2 will elaborate on and further refine this approach, which I call the neutrality triangulation approach because it focuses on evaluating the scope of Section 230 not on whether a particular state law is content-neutral but instead on whether the platform conduct that the state law would regulate is content-neutral (facially or as applied). Part IV will apply this neutrality triangulation approach to the claims in *In re Social Media Addiction* in order to provide guid-

---

[122] Soc. Media Cases, No. JCCP 5255, Lead Case No. 22STCV21355, 2023 WL 6847378, at *31 (Cal. Super. Ct. L.A. County Oct. 13, 2023) ("The features themselves allegedly operate to addict and harm minor users of the platforms regardless of the particular third-party content viewed by the minor user.").

ance to courts and state lawmakers on when, based on this approach, Section 230 permits regulation of addictive design—and when it does not.

1. *In re Zoom*

In re Zoom Video Communications, Inc. Privacy Litigation (*In re Zoom* for short) was a putative class action brought by users of the "Zoom" video conferencing platform against the maker.[123] The plaintiffs were victims of "Zoom bombing," *i.e.*, unwanted intrusions by strangers into their supposed-to-be private video conferences. Some unwanted intruders shared particular content that harmed other users (including church and school groups) in various ways, such as child pornography, hateful and slur-ridden tirades, and exposure.[124] All the intruders disrupted plaintiffs' meetings.

Fourteen Zoom bombing actions were consolidated on May 28, 2020 as *In re Zoom* in front of Judge Koh of the District Court for the Northern District of California. (Judge Koh has since been appointed to the Ninth Circuit.) The plaintiffs alleged a bevy of claims against Zoom for its role facilitating—and failing to stop or warn them about—Zoom bombing, including "invasion of privacy in violation of California common law and the California Constitution, negligence, breach of implied contract, breach of implied covenant of good faith and fair dealing, unjust enrichment/quasi-contract, violation of the California Unfair Competition Law, violation of the Comprehensive Data Access and Fraud Act, and deceit by concealment."[125] They sought certification of two classes: one nationwide class of all persons who used Zoom, another "minor" class of all persons under the age of 13 who used Zoom.[126]

Zoom moved to dismiss, arguing that the plaintiffs' claims were barred by Section 230. Judge Koh analyzed Zoom's motion using the Ninth Circuit's three-element *Barnes* test. In *Barnes*, the Ninth Circuit had held that Section 230(c)(1) "only protects from liability (1) a provider or user of an interactive computer service (2) whom a plaintiff seeks to treat . . . as a publisher or speaker (3) of infor-

---

[123] *In re* Zoom Video Commc'ns, Inc. Priv. Litig., 525 F. Supp. 3d 1017 (N.D. Cal. 2021).

[124] *Id.* at 1025–26 (describing harms).

[125] *Id.* at 1024.

[126] *Id.*

mation provided by another information content provider."[127] Plaintiffs did not dispute that "the allegedly harmful content at issue was posted by third parties and that Zoom played no role in authoring it,"[128] so Judge Koh found the third element was nominally satisfied. Moreover, Zoom was an "interactive computer service," especially given that courts "interpret the term 'interactive computer service' expansively."[129] That satisfied the first element.

The second element of *Barnes*, however, asks whether a claim "seeks to treat" a service "as a publisher or speaker." On this question, the plaintiffs in *In re Zoom* argued that they sought "to hold Zoom accountable for its failure to provide promised security and privacy during Zoom calls," and that Zoom was not a "publisher or speaker" for these purposes.[130] Like the platforms in *In re Social Media Addiction*, Zoom framed this as an effort to plead around Section 230, arguing that "courts routinely reject such attempts to skirt Section 230."[131]

Judge Koh applied a nuanced test, finding some but not all of the plaintiffs' claims barred by Section 230.[132] Specifically, Judge Koh explained that Section 230 bars claims that "(1) challenge the harmfulness of 'content provided by another'; and (2) 'derive from the defendant's status or conduct as a publisher or speaker' of that content."[133] Section 230 does not bar claims, however, that "are content-neutral" or "do not derive from defendant's status as a publisher or speaker,"[134] such as the Zoom-bombing victims' breach of contract claims. "In sum, section 230 . . . immunizes liability deriving from *moderation* of third-party content. . . . Conversely, section 230(c)(1) *allows* claims that . . . are content-neutral."[135]

---

[127] Barnes v. Yahoo!, Inc., 570 F.3d 1096, 1100 (9th Cir. 2009).

[128] *In re Zoom*, 525 F. Supp. 3d at 1029.

[129] Dyroff v. Ultimate Software Grp., Inc., 934 F.3d 1093, 1097 (9th Cir. 2019).

[130] *In re Zoom*, 525 F. Supp. 3d at 1030.

[131] *Id.*

[132] *Id.* ("The Court agrees with Zoom in part. As explained below, Section 230(c)(1) largely bars plaintiffs' claims. For instance, plaintiffs cannot hold Zoom liable for injuries stemming from the heinousness of third-party content.").

[133] *Id.* (quoting Barnes v. Yahoo!, Inc., 570 F.3d 1096, 1102 (9th Cir. 2009)).

[134] *Id.* at 1032.

[135] *Id.* at 1034.

In light of this holding, Judge Koh granted the motion to dismiss in part, denied the motion to dismiss in part, and granted plaintiffs leave to file an amended complaint focusing on content-neutral claims (and excluding claims premised on the harmfulness of content provided by third party users).[136] Zoom settled with the plaintiffs soon thereafter.[137]

Judge Koh based her holding that Section 230 does not bar content-neutral claims on the text of Section 230, its legislative history, and case law. As for text, she saw three textual indicia that Section 230(c)(1) is focused on protecting content-based conduct by platforms and other providers of interactive communications services (*i.e.*, "content moderation," especially "blocking and screening of offensive material"). First, the caption of Section 230(c)(1) reads "Protection for 'Good Samaritan' blocking and screening of offensive material," focusing on "affirmative, good-faith acts."[138] Second, Section 230(c)(1) itself protects providers from liability for being "treated as the publisher or speaker of any information provided by another information content provider," ensuring that "an interactive computer service can moderate third-party content without fear that it will be treated as the publisher or speaker" of that content.[139] Third, Section 230's declaration of policy focuses, again, on "encourag[ing] content moderation" and "maximizing user control over what information is received,"[140] and reading Section 230 to permit content-neutral claims such as the invasion-of-security claims based on Zoom bombing would not be inconsistent with the former goal, and it would advance the latter.

Judge Koh might also have noted, as an additional textual argument, that Section 230(c)(1) is written in definite and singular terms, not general or plural terms. It provides that a platform may not be treated as the "publisher or speaker of any *information* provided by *another* information content provider." The terms

---

[136] *Id.* at 1048.

[137] *In re: Zoom Video Communications, Inc. Privacy Litigation*, https://www.zoommeetingsclassaction.com/ (settlement website); Maya Yang, *Zoom Agrees to 'Historic' $85m Payout for Graphic Zoombombing Claims*, GUARDIAN (Apr. 23, 2022, 1:00 PM).

[138] *In re Zoom*, 525 F. Supp. 3d at 1031; *see also* Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157, 1164 (9th Cir. 2008) (en banc) ("the substance of section 230(c) can and should be interpreted consistent with its caption").

[139] *In re Zoom*, 525 F. Supp. 3d at 1031.

[140] *Id.* (citing Section 230(b)(4)).

"publisher or speaker," "any information," and "another information content provider" are not independent or abstract in this sentence. The "information" involved must be information "provided by another information content provider," and the claim must seek to treat the service as a "publisher or speaker" not in the abstract, but as a publisher or speaker of information that was provided by an information content provider. All of this focuses on how a platform might (or might not) discriminate among particular speakers and their expression.

In many Section 230 claims, this distinction makes no difference. Claims premised on true content moderation—blocking certain content, failing to block certain content, or even prioritizing certain content—necessarily entail treating a provider as publisher of particular "information" provided by "another," *i.e.*, the particular information (content) provided by the particular provider (or providers) whose content a platform censored, failed to censor, prioritized, or failed to prioritize. Thus, a claim such as that in *Dyroff*—alleging a service's recommendation connected an adolescent with a drug dealer, leading to their fatal overdose—arises from and has as its premise the particular drug dealer (the content provider) and the content they provided (the drug sale).

The distinction matters when it comes to claims that involve content generally but do not hinge on any particular content (or even discrete category of content), however. The Zoom plaintiffs' security claims alleging that Zoom failed to prevent Zoom bombing are a great example. Third parties and the content they provide were an aspect of that claim—that is what Zoom bombing is, the provision of content by third parties—but the plaintiffs' security claims did not depend on the involvement of any particular third parties, let alone the provision by those third parties of any particular content. They were content related, but not content based, that is, they were content neutral.

If Section 230(c)(1) provided that a platform (or other interactive computer service provider) not be treated as "publisher or speaker of any information provided by *other information content providers*" then its text would apply not just to claims based on particular content, but also to content-neutral claims. That is not, however, what Section 230(c)(1) says—it focuses its protection on "information provided by another information content provider."

The statute's use of the singular in Section 230(c)(1) is particularly notable because other provisions of Section 230 use more general, plural language. For example, Section 230(c)(2)(B) provides that a service shall not be held liable on

account of "any action taken to enable or make available *to information content providers* . . . the technical means to restrict access to material." Here the statute's context makes clear that it is addressing content-neutral tools (tools that would allow third parties themselves to moderate content by others), and the statute uses the broader plural terminology that makes that clear. The change in formulation to the singular in Section 230(c)(1) makes sense, however, given the provision's overarching focus on protecting platforms' choices in moderating particular content.

As for legislative history, Judge Koh pointed to language in the conference report—familiar to readers steeped in Section 230—explaining Congress's intention to protect platforms from liability for "actions to restrict or to enable restriction of access to objectionable online material," including by overruling *Stratton-Oakmont v. Prodigy*.[141] "Missing from the conference report," she explained, "is any intention to immunize conduct unrelated to content moderation, such a [sic] failure to protect users from a security breach."[142]

Finally, as for case law, Judge Koh pointed to Ninth Circuit precedent making clear that Section 230 "does not declare a general immunity from liability deriving from third-party content,"[143] and noted the point from *Barnes* that Section 230(c)(1) protection extends only to "claims that inherently require[] the court to treat the defendant as the 'publisher or speaker' of content provided by another."[144] (Note, again, that *Barnes* refers to a definite "other" provider of content and does not refer generally to "content provided by others" or "of third-party content" generally). She explained that "[c]ontent-neutral claims do not challenge the harmfulness of third-party content . . . [i]t is irrelevant to these claims whether third-party content . . . is good or bad, displayed or hidden."[145]

Judge Koh went on to list three particular cases in which courts had found that Section 230 did not stand as a barrier to content-neutral claims: (1) *HomeAway.com vs. City of Santa Monica*,[146] in which the Ninth Circuit found Section

---

[141] 1995 WL 323710 (N.Y. Sup. Ct. May 24, 1995).

[142] *In re Zoom*, 525 F. Supp. 3d at 1032.

[143] *Id.* at 1032 (quoting Doe v. Internet Brand, Inc., 824 F.3d 846, 852 (9th Cir. 2016)).

[144] *Id.* (quoting Barnes v. Yahoo!, Inc., 570 F.3d 1096, 1105 (9th Cir. 2009)).

[145] *Id.* at 1033.

[146] 918 F.3d 676 (9th Cir. 2019).

230 did not preempt a local ordinance requiring short-term rental platforms refrain from facilitating bookings for unlicensed properties (given that the platforms "face[d] no liability for the content of the bookings"[147]); (2) *Nunes v. Twitter*,[148] in which the District Court for the Northern District of California held Section 230 did not preempt a statute forbidding texts sent without the recipients' consent (because while texts are content, unsolicited texts were forbidden 'whether the content . . . is bad or good, harmful or harmless'")[149]; and (3) *Doe v. Internet Brands*,[150] in which the Ninth Circuit held that Section 230 did not foreclose a claim for failure to warn platform users that criminals were browsing the platform to identify victims for a rape scheme (despite the fact that the mechanism of harm for the failure to warn depended entirely on the users' posting of content including their personal details to the site).

The Ninth Circuit decided *Lemmon* shortly after Judge Koh's ruling in *In re Zoom*, but she might have cited the opinion there, too, in describing the inapplicability of Section 230(c)(1) to content-neutral claims. In holding that Section 230 did not preempt the parents' wrongful death and products liability claims related to the Snapchat "Speed Filter" that the decedents were using at the time of their fatal accident, the Ninth Circuit emphasized the content neutrality of the claims. "Those who use the internet [] continue to face the prospect of liability, even for their 'neutral tools,' so long as plaintiffs' claims do not blame them for the content that third parties generate with those tools."[151] Because "the Parents' claim [did] not depend on what messages, if any, a Snapchat user employing the Speed Filter actually sends," it was not barred by Section 230.[152]

2. Unpacking the neutrality triangulation approach to Section 230

Figure 1 displays visually the applicability of Section 230 to various sorts of claims based on the relationship of potentially-regulated platform conduct to user-generated conduct.

---

[147] *Id.* at 684.

[148] 194 F. Supp. 3d 959 (N.D. Cal. 2016).

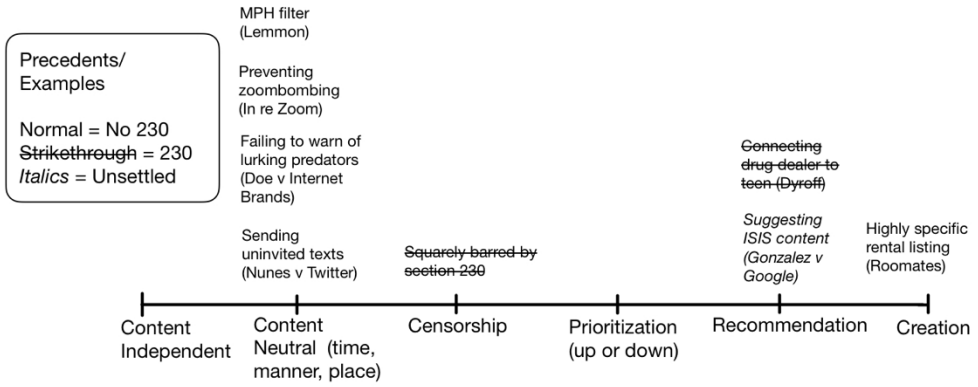
[149] *Id.* at 968.

[150] 824 F.3d 846, 849 (9th Cir. 2016).

[151] Lemmon v. Snap, Inc., 995 F.3d 1085, 1094 (9th Cir. 2021).

[152] *Id.*

Figure 1: Applicability of Section 230 Based on Relationship of Potentially-Regulated Platform Conduct to User-Generated Content



Relationship of Potentially-Regulated Platform Conduct to User Content

Three clarifications are important to understanding this approach to separating content-related platform (or other provider) conduct that is protected by Section 230 from content-related platform conduct that is not protected by Section 230. First, this approach does not speak to (and should not be confused with) the distinct Section 230 question of whether particular content is third-party content or instead the platform's own content. There is an acknowledged exception to Section 230's reach for content *actually created by the Platform*.[153] Courts have found this exception applicable where a platform so heavily shaped the submissions of third parties that it made sense to see the result as the platform's creation, not just the users', and so beyond the reach of Section 230 for that reason.[154] The fact that challenged platform conduct is "content neutral" sometimes comes up in Section 230 caselaw addressing the scope of this exception, because content-neutral tools cannot (by definition) "create" content themselves. So, courts (such

[153] Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157, 1162–63 (9th Cir. 2008) (en banc) ("A website operator can be both a service provider and a content provider: If it passively displays content that is created entirely by third parties, then it is only a service provider with respect to that content. But as to content that it creates itself, or is 'responsible, in whole or in part' for creating or developing, the website is also a content provider. Thus, a website may be immune from liability for some of the content it displays to the public but be subject to liability for other content.").

[154] *Id.*

as the Ninth Circuit in *Roommates.com*) addressing the (in)applicability of this narrow but important exception sometimes point to the "content neutrality" of platform conduct as evidence that the platform should not be seen as the content's *creator*.[155] This precedent simply does not address the distinct question addressed by Judge Koh, however: When can states regulate platform conduct *vis-à-vis* content that is understood to be user- (not platform-) generated?

Second, and more fundamentally, the focus of this approach is ultimately on the content neutrality *vel non* of platform conduct *vis-à-vis* users. A "content-neutral" claim for Section 230 purposes is one that challenges *conduct by a platform* that is content neutral. A "content-based" claim is one that challenges content-discriminatory platform conduct. Implicit in this distinction is what Professor Balkin calls the free speech "triangle" of the internet era. The key relationship in this context is not the relationship between the government, on the one hand, and we the people, on the other. The key relationship is instead a triangle involving the relationships among the government, the platforms, and the people.[156] Judge Koh's "content-neutrality" test focuses not on whether a cause of action or claim is content neutral as between the government and the ultimate speakers (the people). Judge Koh's "content-neutrality" test focuses on whether a cause of action or claim regulates *platform conduct* that is itself content neutral as between the platform and users. Laws regulating platform conduct that is itself content based are barred by Section 230, and laws regulating platform conduct that is itself content neutral are not. The Article uses the term "neutrality triangulation" to emphasize this distinction.

Third, this neutrality triangulation approach is not necessarily inconsistent with other theories of the limits of Section 230, such as the reading of Section 230

---

[155] *Roommates.com*, 521 F.3d at 1171–72 ("[T]he website provided neutral tools . . . but the website did absolutely nothing to encourage the posting of defamatory content . . . . That is precisely the kind of activity for which Congress intended to grant absolution with the passage of section 230.").

[156] Balkin, *supra* note 13, at 2013 ("The twentieth century featured a dualist or dyadic system of speech regulation. In the dualist model, there are essentially two players: the nation-state on the one hand and the speaker on the other."); *id.* at 2014–15 ("On one corner of the triangle are nation-states, states, municipalities, and supranational organizations like the European Union. On the second corner of the triangle are internet-infrastructure companies. . . . On the third corner of the triangle, at the very bottom, we have speakers and legacy media, including mass-media organizations, protesters, civil-society organizations, hackers, and trolls.").

as inapplicable to certain "matchmaking" conduct by platforms. Courts might theoretically read Section 230 to be inapplicable to *both* state regulation of content-neutral platform activity *and* state regulation of platform activity that is content discriminatory in a way that amounts to a "recommendation" or expression of the platform itself. The former is textually grounded in Section 230's focus on specific content posted by specific third parties, and the latter is (arguably) textually grounded in Section 230's limitation to laws that would treat a platform as a "publisher or speaker" of such specific content posted by specific third parties.

That said, reading Section 230 to be inapplicable to state regulation of content-neutral platform activity alone presents a partial solution to the *Gonzalez* puzzle, by permitting state regulation of significant aspects of the "matchmaking" underlying that theory. In oral arguments on *Gonzalez*, justices repeatedly expressed concern about the difficulty of drawing a line between protected and unprotected platform prioritization of content.[157] Any platform (or newspaper for that matter) must decide what content to put at the "top" somehow, even if it simply puts the newest content first. Assuming such prioritization choices must be protected, how could algorithmic recommendations be meaningfully differentiated from such choices—and what line, grounded in the text of Section 230, would differentiate permissible prioritization from impermissible prioritization?

The platforms in *Gonzalez* compellingly argued that in the absence of a clear line companies would respond by acting like all their conduct was potentially subject to liability—defeating the purpose of Section 230.[158] This line-drawing problem is minimized if not eliminated by the neutrality approach, however, because it draws from a well-developed (if occasionally still controversial at the margins) body of caselaw dividing content-neutral regulation, on the one hand, from content-based regulation, on the other. For example, when it comes to state authority

---

[157] *See, e.g.*, *supra* note 86.

[158] Brief for Respondent at 53, 2022 WL 18358194, Gonzalez v. Google LLC, 598 U.S. 617 (2023) ("In a world where websites are pressured to preemptively remove third-party content that might trigger litigation, websites would be even more leery of permitting political (including conservative-leaning) speech on hot-button topics. At the same time, other websites with fewer resources or less public or advertiser pressure might veer in the opposite direction: avoiding liability by refusing to sort, filter, or take down any content."); Leslie Y. Garfield Tenzer & Hayley Margulis, *A 180 on Section 230: State Efforts to Erode Social Media Immunity*, 49 PEPP. L. REV. 49, 78 (2022) ("Stripping platforms of immunity will have a chilling effect on their growth. Platforms will squash speech out of a fear of lawsuits.").

to regulate prioritization, such as the recommendation of ISIS videos at issue in *Gonzalez*, forbidding or deterring recommendations of particular content (like pro-terrorism videos) would not be "content neutral." But forbidding or deterring *personalization* as a tool of prioritization, or particular personalization approaches, like using individualized data drawn from a users' bank statements or online travels without the users' informed consent, would be. Such regulation, by targeting the "learn about you to tailor recommendations" side of Judge Katzman's "matchmaking" hypothetical without any regard to content but leaving untouched the "recommendation" side, would avoid Section 230 preemption under the neutrality triangulation approach.

3. Section 230 as State Action for Platforms

The neutrality triangulation approach to the scope of Section 230 is consistent with Section 230's text, purpose, and circuit precedent, and it also provides a workable answer to the *Gonzalez* problem. More fundamentally, the approach also aligns with underlying constitutional values.

Section 230 is not just any statute, and its interpretation is not a run of the mill question of statutory interpretation. Rather, Section 230 and its interpretation implicate fundamental questions about the future of the freedom of speech, given the significant role that platforms have come to play in constructing the "public" sphere.

From this perspective of Section 230's fit with constitutional values, the neutrality triangulation approach makes sense as a context-specific (and legislatively changeable) "fix" for underlying limitations of the state action doctrine. If Section 230 did not prevent states from regulating content discrimination by platforms, then states could make an end-run around the First Amendment. Where states could not directly engage in content-based regulation themselves (because they are prevented by the First Amendment), they could indirectly engage in content-based regulation by regulating platforms' content moderation choices, including through courts' application of generally applicable laws to particular categories of speech (such as terrorism advocacy, pornography, and so on). For example, where a state could not itself censor speech about terrorism or violence, it (or its courts) could prompt platforms to do so through laws holding them liable for failing to do the same. And where a state could not itself silence particular dissidents, it (or its courts) could prompt platforms to do so through laws holding them liable, one way or the other, for failing to do the same.

In theory, such laundering of content-based state regulation might be checked by the state action doctrine, *i.e.*, by holding that the content-discriminatory actions of a platform that are themselves prompted by a state regulation are acts of the state and so subject to First Amendment protections. The problem, however, is that courts have failed to develop a coherent, muscular state action doctrine capable of such an intervention (though perhaps they will).[159] Part of the challenge is the difficulty of identifying any third-party conduct that is not in some sense influenced by state regulation, let alone drawing lines to identify when state influence is "too much."

This state action challenge is complicated by the fact that First Amendment doctrine must speak in constitutional terms, setting precedents that apply across domains and contexts. Platforms today play an outsized role in shaping public discourse, so state laws have a far greater potential to indirectly (but problematically) regulate the content of public discourse when they are applied to platforms than when they are applied to other sorts of entities. But it would be very hard for constitutional law to develop a state action doctrine capable of calibrating its stringency based on such contextual distinction among regulated actors—such as by requiring one hard-to-trigger "state action" test for regulations of energy companies or day care centers and another more-easily-triggered test for regulations of platforms.

In this context, it makes great sense for Congress, legislatively, to build upon the underlying constitutional framework by imposing a more expression-protective state action doctrine in a particular context. Unlike courts elaborating constitutional doctrine, Congress can invent categories, specify contexts, articulate exceptions, and alter all of these over time as needs vary. And Congress can do so *ex ante*—when a new arrangement is in its infancy—whereas courts can intervene to sculpt state action doctrine only *ex post*, when arrangements may have solidified making intervention too late.

---

[159] *See, e.g.*, Jordon Goodson, *The State of the State Action Doctrine: A Search for Accountability*, 37 TOURO L. REV. 151 (2021) ("The state action doctrine is notoriously confusing and contradictory."). Recent cases have pioneered more aggressive approaches to state action theories in the platform context. *See* Missouri v. Biden, 83 F.4th 350, 373 (5th Cir. 2023) (addressing argument that government coerced platforms to remove content or censor content in violation of the First Amendment), *cert. granted*, 144 S. Ct. 7 (2023).

Understood as a presumptive barrier to state regulation of platforms' content-based activities, Section 230 fills precisely this role. As the law's statement of purpose indicates, it keeps states out of platforms' content moderation choices[160]—though Congress remains free to articulate exceptions. On this understanding, moreover, there is not the same need to apply Section 230's limitations on state authority beyond the context of platforms' content-based activities. States can impose content-neutral regulations on the time, place, and manner of expression themselves, so there is much less risk they would seek to impose content-neutral controls indirectly (by regulating platforms) that they could not impose directly (by regulating users).

III. Charting the First Amendment

Of course, since the first half of the Twentieth Century state regulatory authority has been limited not only by federal preemption (such as through Section 230), but by the Bill of Rights as incorporated against the states by a substantive reading of the Due Process Clause of the Fourteenth Amendment. Thus, the platforms argue that, separate from Section 230, the freedom of speech clause of the First Amendment strictly limits state authority to regulate addictive design. As with Section 230, the resolution of this question depends in part on the resolution of larger, unsettled legal questions—in particular, the question whether and when content moderation is expressive activity protected by the freedom of speech—discussed in subpart A. But, also as with Section 230, there is a strong argument—evident in Judge Kuhl's *Social Media Cases* opinion—that core aspects of addictive design fall outside the reach of the First Amendment regardless of the resolution of those larger questions, as discussed in subpart B. Finally, even state regulation that is subject to the First Amendment may still be permissible if it is sufficiently tailored to advance a content-neutral substantial state interest; subpart C describes such interests implicated by addictive design.

A. Is Content Moderation Expressive?

Unlike Section 230, the First Amendment only limits state regulation of platform conduct that counts as "expression" for purposes of constitutional law.[161] As a preliminary matter, it is far from clear that the First Amendment even applies to

---

[160] 47 U.S.C. § 230(b).

[161] Enrique Armijo, *Reasonableness as Censorship: Section 230 Reform, Content Moderation, and the First Amendment*, 73 Fla. L. Rev. 1199, 1240 (2021).

content moderation choices by platforms in the first instance. There is, at this writing, a split between the Fifth and Eleventh Circuits about whether platform decisions related to content moderation count as expression.[162] This question has arisen in the context of state laws regulating platforms' content moderation activities directly, such as by forbidding platforms from discriminating on the basis of a user's viewpoint.

In *Netchoice v. Attorney General*, the Eleventh Circuit held that platforms' content moderation choices may be "exercises of editorial judgment" that are expressive and so subject to First Amendment protections.[163] But in *Netchoice v. Paxton*, the Fifth Circuit created an apparent conflict with the Eleventh Circuit by holding that platforms' decisions regarding which content to censor are *not* subject to First Amendment protections.[164] The Supreme Court has taken up the case, and will decide it in the near future.[165]

If content moderation choices are not expressive in the first instance, then there seems little room for First Amendment coverage of addictive design. In such a case, only state regulations triggering the First Amendment for other reasons—such as by compelling speech—would trigger constitutional scrutiny.[166] This Article does not take a position on this larger First Amendment debate, which has been ably covered elsewhere, including in this Journal.[167]

---

[162] *See* Adam Candeub, *Editorial Decision-Making and the First Amendment*, 2 J. FREE SPEECH L. 157, 159–60 (2022) (describing controversy); CONG. RSCH. SERV., LSB10748, FREE SPEECH CHALLENGES TO FLORIDA AND TEXAS SOCIAL MEDIA LAWS 3 (2022), https://crsreports.congress.gov/product/pdf/LSB/LSB10748 (describing cases); NetChoice, LLC v. Att'y Gen., Fla., 34 F.4th 1196, 1230–31 (11th Cir. 2022) (finding platforms' editorial judgment to be protected speech under the First Amendment); *see also* Courtney Kim, *Analyzing the Circuit Split over CDA Section 230(E)(2): Whether State Protections for the Right of Publicity Should Be Barred,* 96 S. CAL. L. REV. 449 (2022).

[163] *NetChoice, LLC*, 34 F.4th at 1203.

[164] NetChoice, L.L.C. v. Paxton, 49 F.4th 439, 445 (5th Cir. 2022).

[165] NetChoice, LLC v. Paxton, 143 S. Ct. 744 (2023).

[166] *See* Zauderer v. Off. of Disciplinary Couns., 471 U.S. 626, 651 (1985); *infra* Part III.C (discussing state interests that might support compelled speech, as in warnings, related to addictive design).

[167] *See* Candeub, *supra* note 162.

That said, courts (at this point the Supreme Court) adjudicating the question of First Amendment coverage for content moderation should be cognizant of potential impacts on state authority to regulate addictive design. They should also be careful in articulating governing tests not to inadvertently—out of a desire to protect platforms' ability to express a viewpoint through editorial choices—insulate unrelated platform activities, including addictive design.

B. Is Conditioning Content Moderation?

In her early *Social Media Cases* opinion, Judge Kuhl rejected the platforms' global First Amendment objections to addictive design claims based on a strong argument that much addictive design would remain beyond the reach of the First Amendment even if the Supreme Court were to adopt the Eleventh Circuit's "editorial expression" theory, because much addictive design does not entail content moderation.[168] The Eleventh Circuit did not hold that all platform activities are expressive, rather, it said that platforms may "speak through content moderation."[169] And it was clear that by "content moderation" it meant (and said it meant) removing content, declining to remove content, as well as prioritizing or deprioritizing content.[170]

---

[168] Soc. Media Cases, No. JCCP 5255, Lead Case No. 22STCV21355, 2023 WL 6847378, at *37 (Cal. Super. Ct. L.A. County Oct. 13, 2023) ("Defendants fail to demonstrate that the design features of Defendants' applications must be understood at the pleadings stage to be protected speech."); *id.* ("The allegedly addictive and harmful features of Defendants' platforms are alleged to work *regardless* of the third-party content viewed by the users."); *id.* at *38 ("the design features of Defendants' platforms are not an instance of 'content moderation' as discussed in *NetChoice*").

[169] NetChoice, LLC v. Att'y Gen., Fla., 34 F.4th 1196, 1210 (11th Cir. 2022).

[170] *Id.* ("[W]hen a platform removes or deprioritizes a user or post, it makes a judgment about whether and to what extent it will publish information to its users—a judgment rooted in the platform's own views about the sorts of content and viewpoints that are valuable and appropriate for dissemination on its site.); *id.* ("When a platform selectively removes what it perceives to be incendiary political rhetoric, pornographic content, or public health misinformation, it conveys a message and thereby engages in 'speech' within the meaning of the First Amendment."); *see also id.* at 1204 (explaining that platforms "exercise[] editorial judgment" by removing posts and arranging or prioritizing posts); *id.* at 1204–05 ("[T]he platforms invest significant time and resources into editing and organizing—the best word, we think, is *curating*—users' posts into collections of content that they then disseminate to others. By engaging in this content moderation, the platforms . . . promote various values and viewpoints.").

In assessing the coverage of the First Amendment as in assessing the scope of Section 230, then, the distinction between content-based platform activity and content-neutral platform activity (neutrality triangulation) becomes key in establishing a zone of state authority that does not depend on larger, unresolved legal debates. Reading both the text of the Eleventh Circuit's opinion (focused on "content moderation") and its logical premise (that editorial decisions express a viewpoint) to their maximum, the Eleventh Circuit's opinion simply does not apply to content-neutral platform activities. Key to the logic of the Eleventh Circuit in finding that certain content moderation decisions constitute protected expression is the idea that, like a newspapers' editorial page, a platforms' decisions about what content to prioritize (or recommend) constitute a position in and of themselves.[171] (As an aside, the question of First Amendment coverage therefore interacts in important ways with the *Gonzalez*/matchmaking question about the limits of Section 230,[172] though exploring that interaction is beyond the scope of this

---

[171] *Id.* at 1210 ("The Supreme Court has repeatedly held that a private entity's choices about whether, to what extent, and in what manner it will disseminate speech—even speech created by others—constitute 'editorial judgments' protected by the First Amendment.").

[172] Both require the determination of some hard-to-specify point at which a platform's (or other provider's) conduct ceases to merely entail prioritization or curation (that is protected from regulation by Section 230 but not by the First Amendment) and becomes affirmative recommendation, endorsement, or communication (that may not be protected by Section 230 but that is protected as "expression" by the First Amendment). Interestingly, the circuits have punted on exploring this overlap in the leading cases addressing the applicability of the First Amendment to platform conduct: The Fifth Circuit found the Section 230 issue forfeited in its ruling on the First Amendment issue. NetChoice, L.L.C. v. Paxton, 49 F.4th 439, 469 (5th Cir. 2022). The Eleventh Circuit strangely addressed the First Amendment (constitutional) question but *not* the Section 230 (statutory) question in its ruling. *See NetChoice, LLC*, 34 F.4th 1196. This is unusual because courts normally address potentially dispositive statutory issues before addressing constitutional issues. *See generally* Anita S. Krishnakumar, *Passive Avoidance*, 71 STAN. L. REV. 513 (2019).

Courts could construct an interpretation that harmonizes these two questions by employing the neutrality triangulation approach described in this Article. There would be some sense to a regime in which state regulation of platforms' content-based conduct was entirely foreclosed (by Section 230) until some point at which it became regulable with good reason but subject to enhanced judicial scrutiny (via the First Amendment). This would prevent states from regulating content-based expression indirectly by regulating platforms' content-based conduct in most cases, but allow them to do so in a narrowly tailored way in a subset of the most egregious cases in which individualized recommendations cause direct harm, like the connection of an adolescent with a fentanyl dealer in *Dyroff*.

Article.) Indeed, the Eleventh Circuit expressly held in *NetChoice* that one aspect of the challenged law—its user-data-access requirements—did "not trigger First Amendment scrutiny" because the provision did not "prevent or burden to any significant extent the exercise of editorial judgment."[173] In other words, the Eleventh Circuit held that the one aspect of the law that challenged *content-neutral* platform conduct, and did not interfere with the platforms' *content-based* decisions about user conduct, was not protected by the freedom of speech.

Thus, in *In re Social Media Addiction* the plaintiffs argue that "a slot machine is not a form of speech."[174] This is a powerful argument. If content-neutral platform activities such as the use of intermittent reinforcement and variable reward techniques were held to be "expressive" and protected by the First Amendment, it is hard to say why slot machines—or increasingly digital and increasingly social forms of smartphone-enabled gambling that blur the lines between "social media" and traditional gambling—would not be similarly protected.[175] If courts were to adopt such a theory, the entire field of state gambling regulation—which is historically exempt from First Amendment scrutiny and features a tradition of state regulation long pre-dating the incorporation of the First Amendment as against states[176]—could be upset, especially as it moves into the digital age.[177]

Moreover, there are other strong arguments against reading the First Amendment to protect addictive design through content-neutral techniques such as intermittent reinforcement and variable reward. Unlike content-based features that states have sought to regulate in the past—like violence in video games—the

---

[173] *NetChoice, LLC*, 34 F.4th at 1223.

[174] Pls.' Opp., *supra* note 117, at 24.

[175] *See generally* Nathaniel Meyersohn, *The Dark Side of the Sports Betting Boom*, CNN BUS. (Feb. 10, 2023, 11:48 AM) ("In the past five years, there has been an explosion of online sports betting apps from companies like DraftKings, FanDuel and Caesars."), https://www.cnn.com/2023/02/10/business/online-sports-gambling-addiction/index.html.

[176] *See supra* notes 37–47 and accompanying text.

[177] *Cf.* Commonwealth v. Sadler Brothers Oil Co., No. 230610, 2023 WL 9693656 (Va. Oct. 13, 2023) (questioning whether "skill games" are protected by the First Amendment); *id.* at 11 ("Although at times it is difficult to determine where a particular activity falls on the speech/conduct continuum, no such difficulty is present when the activity being regulated is gambling. We long have viewed gambling as conduct that may be heavily regulated and even banned by the Commonwealth as an exercise of its police powers.").

presence of addictive design is invisible to the user, who may never know they have been exposed to operant conditioning, or may not discover the fact until it is too late. This fact simultaneously vitiates any claim that such techniques are expressive—how can a hidden, unarticulated, and undisclosed pattern of rewards and stimulation be expressive?—and increases the need for state regulation to protect users from unwitting exposure. And finally, addictive design itself interferes with users' liberty—their freedom of thought when subjected to an unwanted and persistent compulsion, and their bodily autonomy when addictive design contributes to mental illness.[178]

C. Are State Interests Content Neutral?

Finally, where the First Amendment applies it does not forbid state regulation altogether; rather, courts apply tests of fit and justification to separate constitutional regulation of speech from unconstitutional regulation. (That said, it is difficult to predictably satisfy courts applying these tests even for legislators working in good faith to address a public concern consistent with constitutional requirements, as recent judicial opinions invalidating state efforts to regulate platforms have made clear.[179])

For any addictive design regulation that courts concluded was subject to the First Amendment, the governing test would likely be intermediate scrutiny (because the regulation was content neutral, because the expression was commercial, or because it entailed a compelled disclosure).[180] Intermediate scrutiny asks

---

[178] Lawrence, *supra* note 22, at 298–99, 300–01, 313–15.

[179] *See, e.g.*, Order Granting Motion for Preliminary Injunction, *NetChoice, LLC v. Bonta*, 2023 WL 6135551 (N.D. Cal. Sept. 18, 2023) (granting preliminary injunction against California's child protection law).

[180] Where the First Amendment is triggered because a regulation compels speech, such as a warning, the appropriate test is technically the *Zauderer* test, which asks whether a disclosure requirement is "reasonably related" to the government interest in "preventing deception of consumers." CONGRESSIONAL RESEARCH SERVICE, ASSESSING COMMERCIAL DISCLOSURE REQUIREMENTS UNDER THE FIRST AMENDMENT (2019) (quoting Zauderer v. Off. of Disciplinary Couns., 471 U.S. 626 (1985)). For protected commercial speech, intermediate scrutiny is provided under the *Central Hudson* framework, which requires that the government must assert a "substantial" interest, the regulation must "directly advance" that interest, and the regulation must be no "more extensive than is necessary to serve that interest." Cent. Hudson Gas & Elec. Corp. v. Pub. Serv. Comm'n, 447 U.S. 557 (1980). For present purposes I focus on the overarching intermediate scrutiny framework, though it is certainly possible that in particular contexts it might be necessary to

whether a state regulation is "narrowly tailored to serve the government's legitimate, content-neutral interests."[181] This can be broken down into three elements: (1) whether the regulation advances a substantial state interest that is content neutral, (2) whether the regulation is narrowly tailored to advance that interest, and (3) whether any less-speech-restrictive alternatives exist.[182]

For present purposes, a key question is what content-neutral, substantial state interests a state regulation of addictive design might advance. Here, courts may well distinguish among methods of addictive design targeted by or motivating a state regulation, either due to First Amendment concerns or due to Section 230 concerns. For example, the Supreme Court has held that states have a substantial interest in protecting children from promotion of illicit drug use.[183] The fact that the Court has recognized as substantial such an arguably content-based interest is yet more evidence of the longstanding tradition of, and respect for, state regulation of addictive products. But there is a real risk that any regulation of platforms narrowly tailored to advance that end would either itself run afoul of Section 230 because it is content based (by impermissibly targeting platforms' decisions about censorship, prioritization, or access to drug-promotion content) or be insufficiently tailored if content neutral (if it sought to avoid Section 230 by regulating platform conduct more generally). (The same would be true of a law that sought to stop gun violence by pressing platforms to discriminate against violent content.[184])

Thus, courts are likely to scrutinize whether professed state interests underlying regulations targeting addictive design are themselves content neutral, and policymakers would be wise to consider this as well in developing reforms. States' interest in protecting their residents' freedom of thought and bodily autonomy,[185]

---

tease out the differences between these particular flavors of intermediate scrutiny. *Cf.* Ashutosh Bhagwat, *The Test That Ate Everything: Intermediate Scrutiny in First Amendment Jurisprudence*, 2007 U. ILL. L. REV. 783 (2007); Aziz Z. Huq, *Tiers of Scrutiny in Enumerated Powers Jurisprudence*, 80 U. CHI. L. REV. 575 (2013).

[181] Ward v. Rock Against Racism, 491 U.S. 781, 798–99 (1989).

[182] *Id.*

[183] *See* Morse v. Frederick, 551 U.S. 393, 408 (2007).

[184] Brown v. Ent. Merch. Ass'n, 564 U.S. 786, 791 (2011).

[185] *See* FARAHANY, *supra* note 30, at 166 ("[W]hen a person or entity tries to override our will by making it exceedingly difficult to act consistently with our desires, and they act with the inten-

including freedom from addiction,[186] is such an interest: it is undermined by addictive design meant to foster compulsion in unwitting users, regardless whether or how particular content might be connected to that compulsion. As I explained in *Addiction and Liberty*, "[u]nderstanding addiction as a deprivation of liberty" supports a *targeted* (and inherently limited) government public health interest in regulating addictive design "because liberty interests protected by the Due Process Clause are an important source of compelling state interests that can justify intrusion on other constitutionally-protected liberties."[187]

Similarly, states' broader interests in protecting the public health,[188] including their residents' mental health, is a content-neutral interest depending on the way in which addictive technology might harm residents. Hypothetically, if harms found to be associated with social media that can broadly be framed as "mental health harms" actually harmed residents' mental health only through "doomscrolling" of bad news, then regulation to protect such harms might well fail neutrality triangulation and so be barred by Section 230 (for regulating platform decisions about user access to bad news). So understood, such harms might also be impermissibly content based so as to not support regulation of conduct because it is protected by the First Amendment,[189] regardless whether the state's interest was technically described at a higher level of generality (like "protecting mental health" or "protecting public health").

These subtleties necessarily would depend on context and require careful attention in individual cases. Indeed, turning back to *In re Social Media Addiction*, the resolution of the claims there, according to the framework just described, depends on the specifics of *how* the platforms allegedly violated state law.

---

tion to cause actual harm, they violate our freedom of action, and our right to cognitive liberty should be invoked as a reason to regulate their conduct.").

[186] *See* Lawrence, *Addiction and Liberty*, *supra* note 22, at 292–93 (arguing States have a compelling interest in protecting residents from being unwittingly exposed to addictive products); Morgan, *supra* note 4 ("Freedom of speech should not include the freedom to inflict a disease.").

[187] Lawrence, *Addiction and Liberty*, *supra* note 22, at 294–95.

[188] Berman, *supra* note 23, at 543 (discussing protection of public health as a state interest potentially supporting regulation of interactive computer service providers).

[189] First Amendment doctrine treats a regulation as content based if discriminating among content was the purpose of the regulation, regardless of whether the regulation does so explicitly. Ward v. Rock Against Racism, 491 U.S. 781, 791 (1989).

IV. Addictive Design and the Content Neutrality Safe Harbor

As the forgoing has explained, the question of state authority to regulate addictive design achieved through content-based platform (or other provider) conduct (like decisions about which content to recommend) is tied up in larger looming legal controversies about the scope of Section 230 and coverage of the First Amendment. But regardless of how courts resolve those controversies, there is a core zone of state authority—a safe harbor—to regulate addictive design achieved through content-neutral platform (or other provider) conduct (like decisions about the sequencing of rewards and interactions through a platform). The remaining question is: Which aspects of current addictive design claims fall in the safe harbor because they regulate content-neutral platform conduct, and so are not barred by Section 230 and the First Amendment regardless of how courts rule on larger content moderation issues?

This Part develops a conceptual roadmap by illustrating the application of the neutrality triangulation approach to certain claims in *In re Social Media Adolescent Addiction Litigation.* Subpart A addresses the question whether the addictive design claims in the case are *inherently* content based, and offers suggestions for states and researchers hoping to develop and inform tailored regulations. Subpart B addresses particular claims.

A. Is Behavioral Addiction Inherently Content Based?

Neither party in *In re Social Media Addiction* cites Judge Koh's opinion in *In re Zoom* or explicitly applies neutrality triangulation to the claims in the case, but many of their arguments—especially as to whether the claims are "content neutral" for purposes of the First Amendment—speak directly to the neutrality triangulation analysis. For the most part, the content neutrality of addictive design claims must be analyzed at the granular level of particular claims and design features, with some likely permitted through neutrality triangulation and others not, as described below. The platforms make one argument that might be understood as asserting that *all* addictive design claims are inherently content based, however.

Specifically, the platforms assert (on page 1 of their motion to dismiss in *In re Social Media Adolescent Addiction*) that "plaintiffs' alleged addiction is to consuming content,"[190] and assert elsewhere that allegedly addictive design features

[190] Defendants' Supplemental Joint Motion to Dismiss Pursuant to Rule 12(b)(6) Plaintiffs' Priority Claims Under Section 230 and the First Amendment at 1, *In re* Soc. Media Adolescent

merely "maximiz[e] user engagement,"[191] by "making certain speech (or speech generally) more engaging, prominent, or interesting."[192] In a similar vein, the platforms assert that the plaintiffs "take advantage of the MDL procedure by pleading claims more generally."[193] And the platforms repeatedly reference a hypothetical that, if their platforms merely displayed endless loops of "blank boxes," then they would not be (allegedly) addictive.[194]

I may be over-reading these passages, but they might be understood as suggesting that individual plaintiff claims, broken down, are ultimately all about consumption of *particular* engaging content—be it beauty comparison, doom scrolling, pornography, political debate, or otherwise—and *not* about either addiction to plaintiffs' products (Instagram, Facebook, Snapchat) themselves, or features of these products (infinite scroll, "likes," etc.).

If this framing of plaintiffs' claims in *In re Social Media Addiction* were correct, then in a sense the general phrases "addictive design," "mental health," and "public health" would be content-neutral covers—a sleight of hand, an obfuscating upward generality shift—for grouping together individually content-based claims (for "political radicalization," "pornography addiction," "body shaming," and the like) into a larger frame as a means to avoid Section 230 and the First Amendment. If plaintiffs could not bring a claim for "exposing kids to violence," "exposing kids to sexual content," or "exposing kids to terrorist content" without violating Section 230 (or triggering the protections of the First Amendment), this theory might have it, then they can't seek to avoid preemption by stitching several such claims together into a larger whole that is no more than the sum of its component parts.

While conceptually possible, the problem with any such argument—which Judge Kuhl pointed out in her *Social Media Cases* opinion[195]—is that it makes

---

Addiction/Pers. Inj. Prods. Liab. Litig., No. 4:22-MD-03047-YGR-TSH (N.D. Cal. Jun. 27, 2023) [hereinafter Platforms' Mot.].

[191] Platforms' Reply, *supra* note 92, at 6.

[192] *Id.* at 12; *see also id.* at 14 ("plaintiffs . . . challenge Defendants' role in making protected third-party speech more available and engaging").

[193] *Id.* at 3.

[194] *See* Platforms' Mot., *supra* note 190, at 1, 18.

[195] Soc. Media Cases, No. JCCP 5255, Lead Case No. 22STCV21355, 2023 WL 6847378, at *39 (Cal. Super. Ct. L.A. County Oct. 13, 2023) ("Defendants are correct that there are allegations in

contested factual assumptions about the interaction of technology and mental health or, at least, about the nature and causes of the plaintiffs' alleged compulsions. The platforms say (without citation) that the "plaintiffs' alleged addiction is to consuming content"—but must this be so? Would one say that a problem slot machine gambler is "addicted to money"? Or are they addicted to slot machines?[196] Do operant conditioning techniques make *content* on the platforms "more engaging," as the defendants assert, or do they make the *platforms themselves* "more engaging"?

While the science in this space is fast-developing (and might develop more quickly with disclosure of the platforms' internal research, as occurred when tobacco litigation reached discovery[197]), operant conditioning research has long focused on *content-neutral* techniques for fostering compulsion, techniques like varying the timing and amount of rewards and pushing repeated, daily interactions.[198] More recently, researchers often focus on "digital addiction" as a category, exploring evidence of addiction to particular devices and platforms, not just to

---

the Master Complaint that could be read to state that Plaintiffs were also harmed by *content* found on Defendants' platforms. But the Master Complaint can be read to state that Plaintiffs' claims are based on the fact that the design features of the platforms—and not the specific content viewed by Plaintiffs—caused Plaintiffs' harms.").

[196] Jonathan Parke & Mark D. Griffiths, *Gambling Addiction and the Evolution of the "Near Miss,"* 12 ADDICTION RSCH. & THEORY 407, 407 (2004) ("There are also multiple stimuli that may be perceived to be rewarding in specific gambling settings because they produce excitement, arousal, and tension *e.g.*, pre-race and race sequence at the race track, the flashing lights of a slot machine, the spinning roulette wheel, the placing of bets. The basic proposition is that gambling behaviour is maintained by winning and losing sequences within an operant conditioning paradigm.").

[197] *See* Engstrom & Rabin, *supra* note 42, at 304 ("The claims were also sufficiently robust to survive pretrial skirmishes. This was crucial, for once the claims survived motions to dismiss, plaintiffs were entitled to discovery—and once discovery commenced, the companies' many secrets spilled out. The resulting picture was devastating. Among other stratagems, the discovery process revealed that the industry had supported research designed to spread disinformation about the hazards of smoking, manipulated cigarettes' nicotine content, and specifically cultivated children, adolescents, and teens as 'replacement' smokers (waiting in the wings, once the current crop of users expired). Documents also underscored the extent to which the industry's public statements, which had for so long denied or minimized the hazards of smoking, were recklessly made and incontrovertibly false.").

[198] *See* U.S. OFF. OF THE SURGEON GEN., *supra* note 32, at 8–9.

particular content.[199] Indeed, in the recent advisory on the harms of social media use for adolescents, the Surgeon General separated the discussion of research into two sections, one addressing harms from "content exposure" and another, distinct section addressing harms from "compulsive or uncontrollable use."[200]

Similarly, lay people explaining the experience of compulsive use of apps and devices often talk about it in content-neutral terms. A simple google search for "addicted to Instagram" or "addicted to TikTok" yields a host of testimonials and DIY treatment guides that frame the undesired, persistent mental affliction they address (whatever name we give it) as technology, not content, based. Relatedly, a simple search in the Apple App Store or Google Play Store reveals numerous products that purport to help users control their technology-related behavioral addictions, and these products enable or disable access to *entire apps*, not just particular categories of content across apps.[201]

To be sure, lay people also sometimes describe the experience of feeling compelled to use apps and devices in content-based terms; a simple google search for "addicted to pornography" yields its own host of testimonials and treatment guides. It would be fair to say that a person who alleged that a platform contributed to their formation of a harmful compulsion to consume a particular type of content by using "content-neutral" tools would have to be very careful, if they hoped to proceed purely on a "content-neutral" theory, not to allow content-based allegations to seep in, either in establishing liability or causation. But it does not follow that all addictive design claims—even claims that a platform fostered a purely content-neutral compulsion to use the platform—are content based, or even that those afflicted by content-focused compulsions cannot bring content-neutral claims (though causation may be more difficult to show in such a case).

The plaintiffs in both *In re Social Media Addiction* and the *Social Media Cases,* for their part, allege compulsions to use *the platforms and their features*, not to

---

[199] *E.g.*, Birgitta Dresp-Langley & Axel Hutt, *Digital Addiction and Sleep*, 5 INT'L J. ENV'T RSCH. & PUB. HEALTH 6910 (2022) (describing "digital addiction" literature and exploring relationship between "digital addiction" and sleep).

[200] U.S. OFF. OF THE SURGEON GEN., *supra* note 32, at 6–12; *see generally* SCHÜLL, *supra* note 31(discussing connections between operant conditioning and the gambling industry).

[201] *E.g.*, FREEDOM, https://freedom.to ("Freedom blocks distracting websites and apps."); *id.* ("Join millions of amazing people who use Freedom to . . . live happier, healthier, and more productive lives.").

consume particular content.[202] Judge Kuhl held that sufficed at the motion to dismiss stage in the *Social Media Cases* to rebut the platforms' effort to frame technology addiction as inherently content based,[203] and the same conclusion seems warranted in *In re Social Media Addiction*.[204] That said, this question will continue to be a key one for determining the scope of state authority to regulate addictive design going forward if courts employ the neutrality triangulation approach—not just in pending cases but also as states contemplate new laws.

Thus, there is a clear takeaway for policymakers and researchers addressing addictive design. States seeking to exercise what authority they have despite Section 230 and the First Amendment should be careful to focus on interests that are not inherently content-based, and they should be sure to focus on differentiating evidence of addictiveness that depends on the particularities of content from content-neutral evidence of addictiveness in exploring the factual basis for regulation and crafting responses. Relatedly, researchers aiming to inform the public and policymakers about the risks of addictive technology should be mindful of neutrality triangulation in their own work to the extent possible, so that it can be best positioned to inform responsive regulation that is consistent with state authority and avoids the need for a lawsuit testing legality (or surviving any such suit). Like the Surgeon General, policymakers and researchers should be careful to separate, to the extent possible, research into the causes and effects of consuming particular types of content from research into the causes and effects of compulsive use more generally, because this distinction may be important in supporting state laws and claims capable of surviving legal challenge.

---

[202] *E.g.*, Compl., *supra* note 18, ¶ 12 (alleging defendants designed products to cause "compulsive use of *their apps*") (emphasis added); *id.* ¶ 87 ("Researchers at UCLA used magnetic resonance imaging to study the brains of teenage girls as they used Instagram. They found that girls' perception of a photo changed depending on the number of likes it had generated. That an image was highly liked—*regardless of its content*—instinctively caused the girls to prefer it.") (emphasis added); *id.* 69, ¶ 238 ("Facebook and Instagram owe their success to their defective design, including their underlying computer code and algorithms, and to Meta's failure to warn plaintiffs and Consortium plaintiffs that the products present serious safety risks. Meta's tortious conduct begins before a user has viewed, let alone posted, a single scrap of content.").

[203] Soc. Media Cases, No. JCCP 5255, Lead Case No. 22STCV21355, 2023 WL 6847378, at *39 (Cal. Super. Ct. L.A. County Oct. 13, 2023).

[204] *See* Barnes v. Yahoo!, Inc., 570 F.3d 1096, 1098 (9th Cir. 2009) (when evaluating defendants' motion, the court must view allegations "in the light most favorable to the plaintiff").

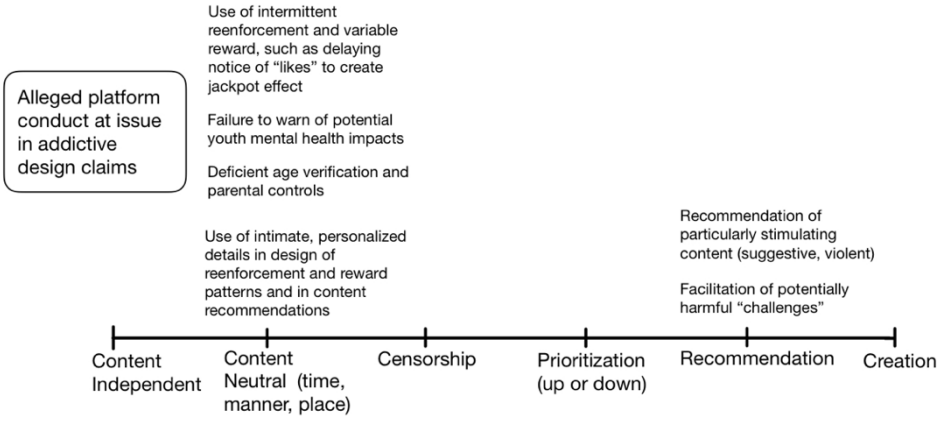
B. Specific Claims

After adopting the neutrality triangulation approach in *In re Zoom*, Judge Koh did not attempt to address the neutrality *vel non* of all of the plaintiffs' claims without the benefit of briefing. Instead, having found that at least some claims were likely content neutral—and at least some were not—she granted Zoom's motion to dismiss in part, denied it in part, and granted plaintiffs leave to amend their complaint to include only content-neutral claims. This approach makes sense in a complicated federal case presenting both a breadth of claims and significant uncertainty about governing legal standards—first clarify the governing standards, then let the parties make their case for how particular claims fare under those standards. (In the *Social Media Cases*, by contrast, Judge Kuhl held that under California procedural rules the legal viability of at least some theories of liability made it unnecessary to address the rest.[205])

[205] *Soc. Media Cases*, 2023 WL 6847378, at *35.

Mindful of the difficulty of working out whether particular claims and conduct in *In re Social Media Addiction* are content neutral without the benefit of the

Figure 2: Relationship of Alleged Platform Conduct at Issue in Addictive Design Claims to User-Generated Content



Relationship of Potentially-Regulated Platform Conduct to User Content

parties' briefing on the particulars, I do not purport here to offer a definitive view on this question. That said, applying the content neutrality approach to the claims and conduct at issue in *In re Social Media Adolescent Addiction*, even in a tentative and impressionistic way, offers an opportunity to illustrate the approach and provide guidance to states seeking to protect their residents from the public health harms alleged in this early case. Figure 2 offers a visual representation.

In their briefs, the platforms concede that some of plaintiffs' addictive design claims challenge "the *manner* in which [the platforms] disseminate and facilitate [] speech,"[206] though they elsewhere assert that the claims are content based in their entirety.[207] The plaintiffs, for their part, insist that their "liability theories are not targeting any particular content or ideas."[208] As they see it, "Defendants could

[206] Platforms' Reply, *supra* note 92, at 1 (emphasis added).

[207] Platforms' Mot., *supra* note 190, at 13 ("No matter how plaintiffs may try to plead around Section 230, the user-generated content on Defendants' services—and its allegedly harmful nature—is the basis for plaintiffs' claims to impose liability on Defendants by treating them as publishers of that content.").

[208] Pls.' Opp., *supra* note 117, at 19.

avoid tort liability by abandoning their unsafe designs, while leaving up the same content."[209]

There are allegations in plaintiffs' complaint that plainly target content-based platform conduct. For example, plaintiffs include numerous allegations that the platforms encourage dangerous "challenges" (such as the "one chip challenge" that recently contributed to the death of a middle school student).[210] A claim premising liability on "encouraging challenges," or a regulation seeking to discourage platforms from recommending or amplifying challenge-related conduct, would regulate content-based platform conduct. Falling outside the content neutrality safe harbor described in this Article, it is difficult to see how such a claim could survive Section 230 unless courts ruled against preemption on the *Gonzalez*/matchmaking issue (and conclude that content moderation is not expressive).

At the same time, plaintiffs also press claims that seemingly target content-neutral platform conduct. Their failure to warn claim alleges, *inter alia*, that the platforms "failed to exercise reasonable care to inform users that . . . [their] products cause addiction, compulsive use, and/or other concomitant physical and mental injuries."[211] Note here that the failure to warn is not about the possibility that users "might be exposed to addictive content," or that users "might be exposed to dangerous challenges"—it is that the products themselves "cause addiction, compulsive use, and/or other concomitant physical and mental injuries." Such a claim raises questions of damages and proof, to be sure, especially in evaluating whether the warning it would require would satisfy applicable common law or statutory standards.[212] But it is seemingly content neutral.

Plaintiffs' allegation that the platforms intentionally made their products difficult to "quit"[213] is also seemingly content neutral, as is there allegation that the

---

[209] *Id.* at 20.

[210] *See* AP, *Maker of the Spicy 'One Chip Challenge' Pulls Product from Store Shelves*, NPR (Sep. 8, 2023, 2:06 AM), https://www.npr.org/2023/09/08/1198369305/maker-one-chip-challenge-pulls-product-from-stores.

[211] Compl., *supra* note 18, ¶ 864.

[212] *Cf.* Chloe Berryman, *Holding Social Media Providers Liable for Acts of Domestic Terrorism*, 72 FLA. L. REV. 1329 (2020).

[213] Compl., *supra* note 18, ¶ 360 ("Even if a user successfully navigates these seven pages, Meta still won't immediately delete their account. Instead, Meta preserves the account for 30 more days. If at any time during those 30 days a user's addictive craving becomes overwhelming and

platforms failed to provide effective parental controls.[214] So, too, are plaintiffs' challenge to particular design choices by the platforms that plaintiffs say made the products more addictive, including infinite scroll and the "loading" wheel. Like regulations requiring newspapers to use recycled newsprint,[215] such claims seem to address content-neutral platform conduct.

There are also, of course, tough claims as to which it is difficult to form even a tentative view. One subset of plaintiffs' claims focus on the platforms' use of algorithmically personalized recommendations in a way potentially analogous to the personalized recommendations of terrorist content challenged in *Gonzalez*. For example, the plaintiffs allege that the platforms push young girls toward content that invites unfavorable beauty comparisons, contributing to problems of self image.[216]

It is hard for the author to form a view from the pleadings of whether such claims are content neutral or not, but the neutrality triangulation approach offers guidance for the questions to ask in determining how Section 230 applies to such claims. The key question going forward is whether these claims are premised on or seek to regulate the *output* of the platforms' processes for prioritizing user content (the content the platforms recommend) or the *input* of those processes (the platforms' use of user-specific information in making recommendations). It is difficult to see how a claim targeting the specific types of content recommended by a platform could be anything other than content based, but, at the same time, a claim targeting the way in which a platform develops recommendations could easily be content neutral.

---

they access the account again, the deletion process starts over. The user must go through all the above steps again, including the 30-day waiting period, if they again wish to delete their account.").

[214] *Id.* ¶ 429 ("Finally, Meta's products offer unreasonably inadequate parental controls; for example, parents cannot monitor their child's account without logging into the child's account directly.").

[215] *E.g.*, CONN. GEN. STAT. ANN. § 22a-256n (West) ("On a state-wide basis, the percentage of recycled fiber contained in newsprint used by all publishers shall be in accordance with the following schedule."); CAL. PUB. RES. CODE § 42760 ("On and after January 1, 1991, every consumer of newsprint in California shall ensure that at least 25 percent of all newsprint used by that consumer of newsprint is made from recycled-content newsprint.").

[216] Compl., *supra* note 18, ¶ 88 (describing "filtered and fake appearances and experiences").

Finally, at points plaintiffs target the "like" button as an addictive design feature. Applying neutrality triangulation, is a platform's inclusion of a "like" button content-based or content-neutral *vis-à-vis* users' expression? This may depend on context, but there is a strong argument that such a feature is content based *vis-à-vis* users' expression (and so outside the neutrality triangulation safe harbor), because it makes it easier for users to communicate a certain viewpoint (approval, or whatever is communicated by a "like"). A platform can to a significant degree control what users communicate about by making it easier to express some things than others—including a "like" button makes it easy to express approval, adding a "thumbs down" makes it easy to express disapproval as well, adding a "share" button makes it easy to communicate the content to others, and so on. On this view, the platform's provision of a "like" button can be analogized to a state's provision of a soap box that could be used only to share positive achievements—which restriction would obviously be content based. That said, some of plaintiffs' allegations related to the "like" button target aspects unrelated to the content-discriminatory function of the feature, such as plaintiffs' claim that likes are tailored to inherently foster a conditioning response and that some platforms intentionally and artificially stagger notice of "likes" received to create a "jackpot" effect or otherwise manipulate the user.[217] Such activity is seemingly content neutral, and so would fall within the safe harbor described here even if the "like" button itself would not.

Conclusion

The addictive potential of new interactive technologies—the potential for smartphones to serve as pocket-sized social slot machines—puts two regulatory paradigms in conflict. The public health paradigm embraces federalism, with states traditionally playing a primary role in protecting their residents from addictive products. But the internet paradigm embraces the market, with states strictly limited in the extent to which they may regulate information content providers. This Article opened with the question of whether these two apparently conflicting approaches can be reconciled when it comes to the public health challenge of addictive design by platforms.

---

[217] *Id.* ¶ 79 ("Instagram's notification algorithm will at times determine that a particular user's engagement will be maximized if the app *withholds 'Likes' on their posts* and then later delivers them in a large burst of notifications.").

While surveying how addictive design fits in the broader Section 230 and First Amendment legal landscape, this Article has also highlighted a "safe harbor" in which states have authority to regulate addictive design regardless of the outcome of broader legal battles. This approach, already employed by Judge Kuhl in the *Social Media Cases*, offers a means of partially reconciling the public health and internet paradigms when it comes to addictive design. Reading Section 230 to forbid states from regulating platforms' content-based conduct but permit states to regulate platforms' content-neutral conduct (and noting that content-neutral conduct is not expressive and so not protected by the First Amendment) preserves a meaningful space for regulation of addictive design even if courts adopt platform-protective answers to larger looming debates about the scope of Section 230 and the First Amendment.

There is an important takeaway from this for nascent efforts to address public health concerns surrounding addictive design. Yes, of course, pay close attention to larger pending legal fights as their outcomes may well influence the scope of state authority in this space. But, at the same time, in order to maximize the legal viability of state regulation in this space, legislators, researchers, and courts should be careful to distinguish between content-based compulsions and conditioning techniques, on the one hand, and content-neutral compulsion and conditioning techniques, on the other.

The argument here has been largely legal, based in the text, history, and purpose of Section 230 as well as First Amendment precedent and values. That said, allow me to conclude by putting my policy views on the table to explain why I have been motivated to highlight arguments supporting the neutrality triangulation approach as a "safe harbor" for some (albeit not all) state regulation of addictive design.

I personally do not find it possible to say for sure that the goal of ensuring innovation on the internet is so important (or the public health threat posed by addictive design so illusory) that open legal questions should all be resolved in favor of maximal preemption of addictive design regulation—that is, that the internet paradigm should govern, completely. At the same time, I am not sure that the public health paradigm should trump outright,[218] either, especially because I take

[218] For a comprehensive and insightful argument in favor of regulation of addictive design and a vision for navigating a path to effective regulation of this sort, see BERNSTEIN, *supra* note 4.

very seriously the anti-tyranny value underlying the First Amendment and (as I explained in Part II.C.3) believe Section 230's insulation of content-based decisions from state regulation can helpfully serve as a sort of enhanced state action doctrine for the distinctive platform and provider context. But while I am not confident that one paradigm should always trump, I am concerned enough by the danger that addictive design poses to both public health *and* innovation online[219]—as well as the danger it might come to pose in a digital future we can only imagine—to feel strongly that we the people, through the laws of our states, should at the very least have the ability to attempt to safeguard ourselves and our kids from unwitting exposure to addictive design where our doing so does not risk an end-run around the First Amendment.[220] The safe harbor for state regulation of content-neutral addictive design described here captures this sweet spot.

---

[219] *See supra* notes 28–29 and accompanying text.

[220] My focus here has been on state regulation, but I'm cognizant that there could be uniformity benefits to federal legislation—especially once the best approach has been developed through state experimentation. Although Congress can theoretically supersede Section 230, a broad reading of Section 230 (or the Constitution) to bar regulation of addictive design would significantly reduce the likelihood of such congressional intervention. There are numerous veto gates through which powerful players (like platforms) can block legislative change and preserve the status quo, and the constitutional dice are heavily loaded against such change. *See* Jonathan S. Gould & David E. Pozen, *Structural Biases in Structural Constitutional Law*, 97 NYU L. REV. 59 (2022). If states are permitted to regulate, then platforms would have an incentive to *support* uniform federal legislation as an alternative to a patchwork of state protections. *See* Nash, *supra* note 45. If, on the other hand, states are not permitted to regulate (and Section 230 is read as effectively a null preemptive measure for addictive design online), then the platforms' incentive would be to oppose any federal measure and preserve a regulatory vacuum.