Vanderbilt Journal of Entertainment & Technology Law

Volume 21
Issue 1 *Issue 1 - Fall 2018*

Article 1

2018

Evil Nudges

Michal Lavi

Evil Nudges

*Michal Lavi**

ABSTRACT

The seminal book Nudge *by Richard Thaler and Cass Sunstein demonstrates that policy makers can prod behavioral changes. A nudge is "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives." This type of strategy, and the notion of libertarian paternalism at its base, prompted discussions and objections. Academic literature tends to focus on the positive potential of nudges and neglects to address libertarian paternalism that does not promote the welfare of individuals and third parties, but rather infringes on it—a concept this Article refers to as "evil nudges." This kind of choice architecture, which negatively influences individual behavior, raises a variety of legal questions and challenges that policy makers must address; yet it remains under-conceptualized.*

Should the law recognize liability for evil nudges that result in bad faith influence? This Article aims to answer this question. It suggests the inclusion of nudges within tort law, arguing that nudges can and should be subject to third-party liability. The inclusion of evil nudges within tort law can be explored broadly, but this Article focuses on one particular case study: the liability of online intermediaries for speech torts caused by evil nudges. This case study provides a natural starting point for considering liability for evil nudges, as designing effective nudges is much easier in a technologically connected environment than in a brick-and-mortar world.

This Article demonstrates that online intermediaries are not just passive middlemen. They influence decisions through website design and promote behavioral change among internet users. The use of big

---

* Ph.D., Cheshin Post-Doctoral Fellow, Hebrew University of Jerusalem-Faculty of Law & Law and Cyber Program, HUJI Cyber Security Research Center. I am most grateful to Professor Tal Zarsky for his insightful comments and guidance in previous stages. I further thank Jonathan Lewy for his valuable input and the Cheshin Fellowship for the financial support. Last but not the least, I thank Alyx E. Eva, Erin E. Meyers, and their colleagues on the *Vanderbilt Journal of Entertainment & Technology Law* staff for the thoughtful comments, suggestions, and outstanding editorial work.

data, the use of artificial intelligence, and the growing use of Internet of Things (IoT) technologies enables unprecedented hyperinfluence. Drawing on network theory, psychology, marketing, and information systems, this Article further demonstrates how nudges influence the process of information diffusion in digital networks. It shows that by nudging, intermediaries can amplify the severity of speech-related harm.

This Article introduces an innovative taxonomy of nudges that online intermediaries utilize, and explains how nudges influence, change internet users' interactions, and form social relations. Afterwards, it examines case law and normative considerations regarding the liability of intermediaries for choice architecture. It argues that the law should respond to "evil nudges," and it proposes nuanced differential guidelines for deciding cases of intermediary liability. It does so while accounting for basic principles of tort law, as well as freedom of speech, reputation, fairness, efficiency, and the importance of promoting innovation.

TABLE OF CONTENTS

I. INTRODUCTION

The seminal book *Nudge* by Richard Thaler and Cass Sunstein demonstrates that policy makers can prod behavioral changes.[1] A nudge is "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives."[2] According to this notion of "libertarian paternalism," the person who organizes the environment in which people make decisions—a choice architect[3]—may predict individuals' prospective behavior and influence them to act in a preferred way. Nevertheless, individuals are free to opt out.[4] Hence, freedom of choice is preserved.

Thaler and Sunstein's conceptualization of nudging prompted discussions and controversies.[5] Academic literature focuses on the positive potential of nudges, but shies away from discussing libertarian paternalism that does not promote welfare but rather infringes on it: "evil nudges."[6] This Article strives to fill in this gap and examines the choice architecture that facilitates defamatory content.[7] Such choice architecture negatively influences individuals' behavior, pushing them to disseminate defamatory speech and exacerbate reputational harm.

This Article focuses on a particular case study of online intermediary liability—namely, website operators that offer platforms for users to create their own content, such as review websites, blogs, discussion forums, and social networks. Intermediaries use advanced technologies to structure the flow of information and interactions. Moreover, intermediaries influence speech by incorporating nudges into the twenty-first century's hyperconnected environment.[8] Consider the following examples:

1. An online intermediary operates a popular website titled The Dirty;[9]

---

1. *See* RICHARD H. THALER & CASS R. SUNSTEIN, NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS 6 (2008).
2. *Id.*
3. *See* RICHARD H. THALER ET AL., CHOICE ARCHITECTURE 1 (2010).
4. *See* THALER & SUNSTEIN, *supra* note 1, at 5–6.
5. *See infra* Section I.A; *see also infra* text accompanying notes 33, 35–36.
6. *See infra* Section I.A, I.D (expanding on "evil nudges").
7. *See* THALER & SUNSTEIN, *supra* note 1, at 10; *infra* Section II.A.
8. *See infra* Section II.C. Intermediaries use advanced technologies to enhance their influence. *See infra* note 19 and accompanying text.
9. *See* THE DIRTY, http://thedirty.com/#F28L2h4Hpb98IDvS.99 [https://perma.cc/39DJ-P9ND] (last visited Sept. 25, 2018); Kate Knibbs, *Cleaning Up the Dirty*, RINGER (Apr. 19, 2017), https://www.theringer.com/2017/4/19/16041942/the-dirty-nik-richie-gossip-site-relaunch-4a086aa24536 [https://perma.cc/HY8J-JVPQ]. TheDirty.com publishes content that contains

2. An online intermediary operates a review website and requires users to categorize their reviews. Most of the categories offered are negative, such as "rip off," "con artists," and "corrupt companies";[10]
3. An online intermediary encourages users to publish rumors, gossip, and defamatory content by using slogans such as "Keep it juicy."[11] Some of the encouragements are general, while others are personalized and adjusted to users' characteristics.[12]

Given these intermediaries' choice architecture, defamation and negative fake stories unsurprisingly fill their platforms. One potential solution would be to allow victims of offensive speech to file libel suits against intermediaries that facilitate offensive and harmful content published on their sites. Should the law impose liability on intermediaries for acting in bad faith by enhancing offensive speech through nudges? How should the courts treat such nudges? Finally, which standards of liability should be set?

Technological design organizes the world for us—subtly shaping the ways that we make sense of it.[13] Every choice a web designer makes

---

rumors, speculation, assumptions, opinions, and factual information. Postings may contain erroneous or inaccurate information. *See id.*

10. *See* Glob. Royalties, Ltd. v. Xcentric Ventures, LLC, 544 F. Supp. 2d 929, 932 (D. Ariz. 2008).

11. *See* Nancy S. Kim, *Website Design and Liability*, 52 JURIMETRICS 383, 393 (2012). This was the slogan of the intermediary of JuicyCampus.com. *See* Associated Press, *Lawsuits, Weak Economy Kill JuicyCampus.com*, FOX NEWS (Feb. 5, 2009), http://www.foxnews.com/story/2009/02/05/lawsuits-weak-economy-kill-juicycampuscom.html [https://perma.cc/Y4E8-U48B]. Similarly, GossipReport.com encouraged its users to think about and report on controversial issues. *See* Gene Weingarten, *Lying Liars*, WASH. POST (June 15, 2008), http://www.washingtonpost.com/wpdyn/content/article/2008/06/11/AR2008061103226.html?noredirect=on [https://perma.cc/E6BR-8SRA]. TheDirty.com went further and included a button labeled as "submit dirt." *A Dirty Job: TheDirty.com and Liability for User Content*, LAW360 (June 8, 2012, 12:33 PM) https://www.law360.com/articles/347948/a-dirty-job-thedirty-com-and-liability-for-user-content. The cellular app, Secret (that was recently shutdown), also encouraged distribution of gossip, rumors, and personal information anonymously. *See* Mike Isaac, *A Founder of Secret, the Anonymous Social App, Is Shutting It Down*, N.Y. TIMES (Apr. 29, 2015), http://www.nytimes.com/2015/04/30/technology/a-founder-of-secret-the-anonymous-social-app-shuts-it-down-as-use-declines.html [https://perma.cc/M54Y-MGWC].

12. *See* Kim, *supra* note 11, at 403. Intermediaries can use big data, artificial intelligence and machine learning to personalize encouragements. *See, e.g.*, Dyroff v. Ultimate Software Grp., Inc., No. 17-cv-05359-LB, 2017 WL 5665670, at *21 (N.D. Cal. Nov. 26, 2017).

13. *See* WOODROW HARTZOG, PRIVACY'S BLUEPRINT: THE BATTLE TO CONTROL THE DESIGN OF NEW TECHNOLOGIES 22, 26 (2018) (emphasis in original) (explaining that design is everything since "[d]esign can act as a medium, communicating *on behalf of* both designers and users. It can also act *upon* users, constraining or enabling them in particular ways."); Julie E. Cohen, *What Is Privacy*, 126 HARV. L. REV. 1904, 1913 (2013).

affects users' behavior, interpersonal dynamics,[14] and decisions to generate and disseminate content.[15] The examples above represent three common online strategies of nudges: (1) "Focal Point"; (2) "Channeling & Leading"; and (3) "Encouragement."[16] These examples are not theoretical. In fact, policy makers and courts discuss them regularly.[17] Yet, the scope of intermediary liability for nudges remains unclear. Legal scholars, judges, and legislatures lack a systematic understanding of how evil nudges influence internet users, let alone how the law should respond.[18] This Article aims to meet that challenge. It strives to provide a comprehensive framework for online intermediary liability for speech tort nudges. This framework will entail a nuanced, context-specific analysis that is neutral to technological advances. It can be used by judges and policy makers to promote just and efficient decisions on intermediary liability.

Part II focuses on intermediaries' choice architectures and their effect on social behavior, network dynamics, and diffusion of information. Drawing on network theory, psychology, marketing, and information systems, it provides an innovative taxonomy of three main types of nudges. This taxonomy illustrates nudging strategies and how they exacerbate speech tort and falsehoods' harm. Reconceptualizing the influence of intermediaries is particularly important in the age of big data, artificial intelligence, and Internet of Things (IoT) technologies that enable hyperinfluence on a scale like never before.[19]

---

14. *See* James Grimmelmann, *Saving Facebook*, 94 IOWA L. REV. 1137, 1162 (2009); Samantha L. Millier, Note, *The Facebook Frontier: Responding to the Changing Face of Privacy on the Internet*, 97 KY. L.J. 541, 556 (2008). For example, intermediaries in social networks (such as Facebook), enhance motivation to spread content. *See* Grimmelmann, *supra*, at 1162. They are great at making us feel like we know many people. The pictures, names, and other informal touches make contacts look like well-known friends. *See id.* Thus, we share with them information we would not have shared otherwise. *See id.*

15. *See* B.J. FOGG, PERSUASIVE TECHNOLOGY: USING COMPUTERS TO CHANGE WHAT WE THINK AND DO 5 (Jonathan Grudin et al. eds., 2003); JACOB SILVERMAN, TERMS OF SERVICE: SOCIAL MEDIA AND THE PRICE OF CONSTANT CONNECTION 8 (2015).

16. *See infra* Section I.D.

17. *See, e.g.*, Communications Decency Act of 1996, Pub. L. No. 104-104, 110 Stat. 56 (codified as amended at 47 U.S.C. § 230(c)(1) (2012)); Jeff Kosseff, *The Gradual Erosion of the Law that Shaped the Internet: Section 230's Evolution Over Two Decades*, 18. COLUM. SCI. & TECH. L. REV. 1, 40–41 (2016) (indexing several cases concerning online intermediary immunity); *see also infra* Section II.A.

18. *See infra* Sections II.A, II.C. (demonstrating the incoherency in judicial decision and scholarly work).

19. *See* Jack M. Balkin, *Free Speech Is a Triangle*, COLUM. L. REV. (forthcoming 2018) (manuscript at 67) (explaining that the problem of online intermediaries, which sets them apart from twentieth-century mass media companies, is their dangerous ability to manipulate and breach trust by utilizing personal data); Michael Guihot, Anne F. Matthew & Nicolas P. Suzor, *Nudging Robots: Innovative Solutions to Regulate Artificial Intelligence*, 20 VAND. J. ENT. & TECH. L. 385, 446–48 (2017) (discussing nudging and artificial intelligence). Technologies can enhance the efficiency and efficacy of content through design and automation. *See* Lili Levi, *Real "Fake*

This reconceptualization allows for a more comprehensive analysis of intermediary liability.

Part III explores case law regarding intermediary liability in the United States and the European Union, focusing on the inconsistency of court decisions and differences among jurisdictions. It follows with a discussion of normative considerations from a broader perspective.

Part IV advocates for the recognition of nudges as part of tort law. It suggests that the negative influence caused by evil nudges can and should be subject to third-party liability. The Article offers tailored guidelines for determining liability of intermediaries by using differential standards for structuring judicial discretion and promoting proportional liability and consistency. In doing so, this Article takes into account basic principles of tort law, as well as freedom of speech, reputation, fairness, efficiency, and the importance of innovation.

Evil nudges are a major problem throughout society today. Beyond the implications on individual dignity, evil nudges can influence and change social behavior and infringe on our political security and democracy.[20] The potentially grave consequences of bad faith influence are a wakeup call for the law to respond. This Article constitutes the first sustained examination of the role of evil nudges in tort law and rectifies the gap in legal scholarship on third-party liability.[21] Invaluable insights gleaned from intermediary liability can open up new avenues of analytic inquiry and inspire further discussions on contributory liability for evil nudges in general. Thus, this Article takes the first step towards providing a full-fledged theoretical framework for

---

News" and Fake "Fake News", 16 FIRST AMEND. L. REV. (forthcoming) (manuscript at 20). As research indicates, sophisticated players use data analytics and artificial intelligence to increase the efficiency of their messages. *See id.* at 26. These issues are of particular importance in the wake of the Facebook-Analytica scandal. *See* Alexandra Samuel, *The Shady Data-Gathering Tactics Used by Cambridge Analytica Were an Open Secret to Online Marketers. I Know, Because I Was One*, VERGE (Mar. 25, 2018), https://www.theverge.com/2018/3/25/17161726/facebook-cambridge-analytica-data-online-marketers [https://perma.cc/WG8H-DAG7].

20. *See* FUTURE OF HUMANITY INST., UNIV. OF OXFORD, THE MALICIOUS USE OF ARTIFICIAL INTELLIGENCE: FORECASTING, PREVENTION, AND MITIGATION 6 (2018), https://www.eff.org/files/2018/02/20/malicious_ai_report_final.pdf [https://perma.cc/A6PM-QUEA]. The improved ability of intermediaries and other stakeholders to analyze human behaviors, moods, and beliefs by using big data allows them to apply effective evil nudges and manipulate others. *See id.* This threatens a society's ability to engage in truthful, free, and productive discussions about matters of public importance and legitimately implement broadly just and beneficial policies. *See id.* The Facebook-Analytica scandal serves as a good example. *See* Samuel, *supra* note 19. For expansion on the far reaching technological abilities and Artificial Intelligence in particular to hinder our political security, see FUTURE OF HUMANITY INST., *supra*, at 28–29.

21. *See* Assaf Hamdani, *Gatekeeper Liability*, 77 S. CAL. L. REV. 53, 56–57 (2003) ("[T]he topic of third-party liability has received only scant attention by legal academics. . . . [L]ittle is known about the appropriate scope of third-party liability. Specifically, legal scholarship has little to say about the standard of liability that should apply to third parties."); *infra* Section I.A.

third-party liability and adjusting tort law to meet the challenges of the twenty-first century.

II. NUDGES AND NETWORKS: BETWEEN PSYCHOLOGY AND TECHNOLOGY

A. Why Nudge?

Decision makers do not make choices in a vacuum. Instead, they reach decisions based on cultural and social conditions—some of which are visible, while others remain hidden. Moreover, decisions are usually context sensitive.[22] In *Nudge*,[23] Thaler and Sunstein identify broad ways of changing civic behavior in a predictable way without banning options or significantly changing economic incentives.[24] They call this "libertarian paternalism."[25] Choice architects may predict individuals' prospective behavior and influence them to act in a preferred direction.[26] This strategy can solve problems caused by bounded rationality and bounded self-control.[27] Nevertheless, individuals are free to opt out; hence, keeping their freedom of choice.[28] For example, people tend to stick with the status quo when using default options.[29] Knowledge of this bias allows choice architects to set

---

22. *See* THALER & SUNSTEIN, *supra* note 1, at 3.

23. *See* KENT GREENFIELD, THE MYTH OF CHOICE: PERSONAL RESPONSIBILITY IN A WORLD OF LIMITS 198 (2011). *See generally* THALER & SUNSTEIN, *supra* note 1.

24. *See* THALER & SUNSTEIN, *supra* note 1, at 99, 241. For example, a GPS is a nudge. Default rules and disclosure of relevant information (i.e., about the risks of smoking) also count as nudges. *See id.* at 68; Cass Sunstein, *There's a Backlash Against Nudging – But It Was Never Meant to Solve Every Problem*, GUARDIAN (Apr. 24, 2014, 2:30 PM), https://www.theguardian.com/commentisfree/2014/apr/24/nudge-backlash-free-society-dignity-coercion [https://perma.cc/AZ8X-XDDP]. By contrast, a criminal penalty is not a nudge because it imposes "significant material incentives on people's choices." *See* Cass R. Sunstein, *Do People Like Nudges?*, ADMIN. L. REV. (forthcoming) (manuscript at 2); *accord* CASS R. SUNSTEIN, THE ETHICS OF INFLUENCE: GOVERNMENT IN THE AGE OF BEHAVIORAL SCIENCE 39 (2016).

25. *See* CASS R. SUNSTEIN, WHY NUDGE? THE POLITICS OF LIBERTARIAN PATERNALISM 19, 55–56 (2014) (referring to a continuum of costs of choice, and that, when the costs of choice are burdensome—it is hard paternalism, and whenever the costs are insignificant, it is soft paternalism).

26. *See* Cass R. Sunstein, *The Storrs Lectures: Behavioral Economics and Paternalism*, 122 YALE L.J. 1826, 1834, 1887 (2013); Abbey Stemler, *Regulation 2.0: The Marriage of New Governance and Lex Informatica*, 19 VAND. J. ENT. & TECH L. 87, 105–06 (2016) (explaining how choice architecture regulates the flow of information online).

27. *See* Richard H. Thaler & Cass R. Sunstein, *Libertarian Paternalism Is Not an Oxymoron*, 70 U. CHI. L. REV. 1159, 1184 (2003). On the problem of bounded rationality, see Daniel Kahneman, *Maps of Bounded Rationality: Psychology for Behavioral Economics*, 93 AM. ECON. REV. 1449, 1449 (2003) (explaining that when individuals make decisions, their rationality is limited by systematic biases that separate the choices they make from the optimal beliefs and choices assumed in economic rational-agent models); Herbert A. Simon, *A Behavioral Model of Rational Choice*, 69 Q.J. ECON. 99, 99 (1955).

28. *See* Thaler & Sunstein, *supra* note 27, at 1184.

29. *See* THALER & SUNSTEIN, *supra* note 1, at 8 ("[P]eople have a strong tendency to go along with the status quo or default option.").

the default rules and thus influence users' behavior in their preferred ways—knowing that most people will not deviate from the default choice. Yet the choice architect does not forbid decision makers from deviating away from the default choice, nor does he tax deviations from the default—people remain free to make their own choices. This concept can be useful for preventing self-harm, as well as harm from third parties.[30] Advocates of the nudge approach believe that choice-preserving alternatives are preferable to mandates.[31]

The nudge approach has achieved widespread recognition in public policy making, which has led to reforms.[32] However, it has also attracted controversies, objections, and ethical concerns.[33] Much of the criticism surrounding this approach comes from libertarians, who argue that the idea of "libertarian paternalism" contradicts itself.[34] They view the "guiding" of a person's choices and the elimination thereof as comparable.[35] A debate that contemplates whether nudges are ethical

30. *See* Christopher McCrudden & Jeff King, *The Dark Side of Nudging: The Ethics, Political Economy, and Law of Libertarian Paternalism*, *in* CHOICE ARCHITECTURE IN DEMOCRACIES, EXPLORING THE LEGITIMACY OF NUDGING 67, 81, 93 (Alexandra Kemmerer et al. eds., 2016) (referring to texting while driving and fuel standards as areas where nudging is appropriate). Nudges related to distracted driving and fuel standards are simply concerned with preventing harm to third parties. In other words, Sunstein applies nudges as they relate to the prevention of harm to others. *See id.*; SUNSTEIN, *supra* note 24, at 24–25 (differentiating between nudges that prevent harm to self and nudges that prevent harm to others, yet Sunstein concludes that both nudges can be ethical and promote welfare); SUNSTEIN, *supra* note 25, at 108. However, critics of this philosophy caution the use of nudges, instead of mandates, for preventing harm to others. *See* McCrudden & King, *supra*, at 69.

31. *See* Ryan Calo, *Code, Nudge, or Notice?*, 99 IOWA L. REV. 773, 783 (2014); Cass R. Sunstein, *Nudges vs. Shoves*, 127 HARV. L. REV. F. 210, 210 (2014) (demonstrating that the nudge concept can also be used to avoid causing harm to third parties).

32. *See, e.g.*, RICHARD H. THALER, MISBEHAVING: THE MAKING OF BEHAVIORAL ECONOMICS 331 (2015); Pelle Guldborg Hansen & Andreas Maaløe Jespersen, *Nudge and the Manipulation of Choice: A Framework for the Responsible Use of the Nudge Approach to Behaviour Change in Public Policy*, 4 EUR. J. RISK REG. 3, 4 (2013); McCrudden & King, *supra* note 30, at 86; Sunstein, *supra* note 24 (manuscript at 34). The nudge approach was applied to bring behavioral change in public policy in the United States and Europe. *See* CASS R. SUNSTEIN ET AL., TRUSTING NUDGES? LESSONS FROM AN INTERNATIONAL SURVEY 1, 2–3, 17 (2018) (describing a survey on nudges applied by government and concluding that there is generally a high level of approval for nudges as policy tools across different countries—including Belgium, Denmark, Germany, South Korea, and the United States).

33. *See, e.g.*, SUNSTEIN, *supra* note 25, at 137; Cass R. Sunstein, The Ethics of Nudging 1, 12–13 (Nov. 20, 2014) (preliminary draft), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2526341 [https://perma.cc/BUG2-HBL5] (listing and confronting seven objections to nudges).

34. *See* Heidi M. Hurd, *Fudging Nudging: Why 'Libertarian Paternalism' Is the Contradiction It Claims It's Not*, 14 GEO. J.L. & PUB. POL'Y 703, 734 (2016).

35. *See* Hansen & Jespersen, *supra* note 32, at 9; Gregory Mitchell, *Libertarian Paternalism is an Oxymoron*, 99 NW. U. L. REV. 1245, 1265 (2005) (challenging Thaler and Sunstein's arguments as expressed in Thaler & Sunstein, *supra* note 27); Henry Farrell & Cosma Shalizi, *'Nudge' Policies Are Another Name for Coercion*, NEW SCIENTIST (Nov. 2, 2011), https://www.newscientist.com/article/mg21228376-500-nudge-policies-are-another-name-for-coercion/ [https://perma.cc/YJJ3-F253] (arguing that nudges are paternalistic coercion); Brendan

and whether they manipulate choices has ensued.[36] Due to the inevitability of choice architecture and the fact that there is no such thing as a completely "neutral" design, researchers question the basis for objections to nudges.[37] Other researchers question when it is inappropriate to use nudges[38]—attempting to differentiate between strategies of choice architecture.[39]

Empirical research of views regarding nudges found greater support for transparent nudges that appeal to conscious, deliberative thinking, as opposed to nontransparent nudges that affect subconscious or unconscious processing of information.[40] However, people's views cannot resolve the question of when nudges should be constrained—thus, this question remains unanswered.[41] This Article does not aim to resolve these general questions. Instead, it focuses on evil nudges that do not promote social welfare.[42] These nudges are often driven by illicit motives[43] and may lead to severe harm.[44]

Evil nudges are normatively undesirable, and most people would be unhappy to be their target.[45] Thaler and Sunstein briefly addressed evil nudges and suggested transparency as a solution.[46] The notion of

---

O'Neill, *A Message to the Illiberal Nudge Industry: Push Off*, SPIKED (Nov. 1, 2010), http://www.spiked-online.com/newsite/article/9840#.U99QveN_sl8 [https://perma.cc/P668-E2PG] (claiming that individual choice may not exist due to manipulation caused by this policy).

36. *See* Hansen & Jespersen, *supra* note 32, at 13; John Hasnas, *Some Noodging About Nudging: Four Questions About Libertarian Paternalism*, 14 GEO. J.L. & PUB. POL'Y 645, 653 (2016).

37. *See* THALER & SUNSTEIN, *supra* note 1, at 86 (suggesting that it is pointless to discuss liability for choice architecture because it is unavoidable); Hansen & Jespersen, *supra* note 32, at 8, 10 (distinguishing given contexts that *accidentally* influence behavior from situations involving choice architects who *intentionally* attempt to alter behavior by manipulating such contexts).

38. *See* SUNSTEIN, *supra* note 24, at 15–16 (explaining that nudges do not raise ethical questions when they promote autonomy, dignity, and welfare).

39. *See* Hansen & Jespersen, *supra* note 32, at 13 (differentiating nudges that aim to influence behavior maintained by automatic thinking and nudges that aim to influence reflective thinking).

40. *See* Sunstein, *supra* note 24 (manuscript at 3, 39) (explaining that a statement or an action can be manipulative if it does not sufficiently appeal to people's capacity for reflective and deliberative choice—thus failing to respect people's autonomy); SUNSTEIN ET AL., *supra* note 32, at 18 ("Rather, effective and publicly accepted nudges will more likely be developed with a process that includes early participation of the affected groups, public scrutiny, and deliberation – as well as transparent processes in governmental institutions.").

41. *See* Sunstein, *supra* note 24 (manuscript at 1, 4) ("Evidence about people's views cannot resolve the ethical questions, but in democratic societies (and nondemocratic ones as well), those views will inevitably affect what public officials are willing to do.").

42. *See* THALER, *supra* note 32, at 345. Nudges are merely tools. Therefore, when Thaler signs copies of *Nudge*, he always adds the phrase "nudge for good." *Id.*

43. *See* SUNSTEIN, *supra* note 25, at 159–60; THALER & SUNSTEIN, *supra* note 1, at 240–41; Sunstein, *supra* note 26, at 1898; Sunstein, *supra* note 33, at 6.

44. *See infra* Sections I.C, I.D (outlining a taxonomy of nudge torts and explaining their harm potential).

45. *See* Sunstein, *supra* note 24 (manuscript at 27).

46. *See* THALER & SUNSTEIN, *supra* note 1, at 244.

transparency, however, will not likely guarantee an acceptable public policy.[47] Even highly transparent nudges promote behavioral change,[48] and transparency does not resolve the harm caused by evil nudges to third parties. Thus, transparency alone remains a suboptimal solution for mitigating the problem of evil nudges.

A choice architect cannot avoid influencing decisions and behavior.[49] However, a choice architect often intentionally tries to alter behavior by nudging and attempts to exert his own will over other people's actions. In such cases, the influence of choice architecture is not accidental.[50] Therefore, ignoring evil nudges distorts the concept of responsibility. The following sections will focus on a single case study to demonstrate how evil nudges relate to the liability of online intermediaries for speech torts.

B. Intermediaries, Social Contexts, and Online Nudges

The advent of the internet, mobile phones, and online social networks upgraded our ability to constantly stay in touch with one another. This revolution affords new opportunities to create social ties, share ideas, form communities, and engage in diverse social dynamics anywhere and at any time.[51] Once upon a time, people thought the internet was the harbinger of "disintermediation"—a sovereign-free

47. *See* Sunstein, *supra* note 33, at 9. Many researchers have proved that transparency and disclosure do not fulfill their goals. *See* OMRI BEN-SHAHAR & CARL E. SCHNEIDER, MORE THAN YOU WANTED TO KNOW: THE FAILURE OF MANDATED DISCLOSURE 42 (2014); GEORGE LOEWENSTEIN ET AL., WARNING: YOU ARE ABOUT TO BE NUDGED 1, 3 (2014); Sophie C. Boerman, Karolina Tutaj & Eva A. van Reijmersdal, *The Effects of Brand Placement Disclosures on Skepticism and Brand Memory*, 38 COMM. 127, 142 (2013); Florencia Marotta-Wurgler, *Even More Than You Wanted to Know About the Failures of Disclosure*, 11 JERUSALEM REV. LEGAL STUD. 63, 65 (2015); Zahr K. Said, *Mandated Disclosure in Literary Hybrid Speech*, 88 WASH. L. REV. 419, 458 (2013).

48. *See* SUNSTEIN, *supra* note 25, at 147–48. Examples for transparent nudges are found in graphic warnings. *See* THALER & SUNSTEIN, *supra* note 1, at 68. These nudges usually work on intuitive thinking (System 1) in contrast to the analytic system (System 2). *See* DANIEL KAHNEMAN, THINKING, FAST AND SLOW 237 (2011) (explaining the two systems of thinking: intuitive thinking and deliberative thinking); THALER, *supra* note 32, at 109; Sunstein, *supra* note 24 (manuscript at 24–25, 36) (noting that even when people are informed that they are being nudged, the effect of the nudge is usually not reduced).

49. *See* Hansen & Jespersen, *supra* note 32, at 8.

50. *See id.* at 10.

51. *See* LEE RAINIE & BARRY WELLMAN, NETWORKED: THE NEW SOCIAL OPERATING SYSTEM 270 (2012).

medium controlled from the bottom up by users.[52] Instead, it simply created new media gatekeepers that control the flow of information.[53]

Intermediaries are not just middlemen; they act as centers of power and governors of speech.[54] They shape public discourse[55] and play an essential role in shaping social ties and directing the attention of internet users.[56] Intermediaries structure the flow of information.[57] They influence (1) the nature of social dynamics; (2) the content that a platforms' users create, consume, and share; and (3) the likelihood that users further spread content.[58] To do so, they utilize insights gleaned from sociology, psychology, and management.[59] These insights allow intermediaries to predict cognitive biases and social dynamics, deploy new sociotechnical systems, and influence flows of information efficiently in their attempts to gain more profits.[60]

---

52. John Perry Barlow, *A Declaration of the Independence of Cyberspace*, ELECTRONIC FRONTIER FOUND., https://www.eff.org/cyberspace-independence [https://perma.cc/P4BA-XFT2] (last visited Sept. 26, 2018).

53. Jack M. Balkin, *Old-School/New-School Speech Regulation*, 127 HARV. L. REV. 2296, 2297 (2014); Derek E. Bambauer, *Middlemen*, 65 FLA. L. REV. F. 1, 2 (2013); Stemler, *supra* note 26, at 105–06.

54. *See* Jack M. Balkin, *Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation*, 51 U.C. DAVIS L. REV. 1149, 1184 (2018) ("These companies are the governors of these digital communities, and if you have an account and use the service, you are part of the governed."); Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598, 1603, 1609 (2018) (focusing on content moderation).

55. *See* TARLETON GILLESPIE, CUSTODIANS OF THE INTERNET: PLATFORMS, CONTENT MODERATION AND THE HIDDEN DECISIONS THAT SHAPE SOCIAL MEDIA 23 (2018) ("Platforms may not shape the public discourse by themselves, but they do shape the shape of the public discourse. And they know it.").

56. *See* Seth F. Kreimer, *Censorship by Proxy: The First Amendment, Internet Intermediaries, and the Problem of the Weakest Link*, 155 U. PA. L. REV. 11, 68 (2006).

57. *See* Julie E. Cohen, *Law for the Platform Economy*, 51 U.C. DAVIS L. REV. 133, 148 (2017) ("Massively intermediated, platform-based media infrastructures have reshaped the ways that narratives about reality, value, and reputation are crafted, circulated, and contested."); Stemler, *supra* note 26, at 105–06.

58. *See* Michal Lavi, *Online Intermediaries: With Power Comes Responsibility*, JOLT DIG. (May 11, 2018), https://jolt.law.harvard.edu/digest/online-intermediaries-with-power-comes-responsibility [https://perma.cc/YT78-BB7P]. For example, Facebook conducted research to target emotionally vulnerable and insecure youth. In fact, "Facebook can figure out when people as young as 14 feel 'defeated,' 'overwhelmed,' 'stressed,' 'anxious,' 'nervous,' 'stupid,' 'silly,' 'useless,' and [like] a 'failure.' Such information gathered through a system dubbed sentiment analysis could be used by advertisers to target young Facebook users when they are potentially more vulnerable." *See* Nick Whigham, *Leaked Document Reveals Facebook Conducted Research to Target Emotionally Vulnerable and Insecure Youth*, NEWS.COM.AU (May 1, 2017), http://www.news.com.au/technology/online/social/leaked-document-reveals-facebook-conducted-research-to-target-emotionally-vulnerable-and-insecure-youth/news-story/d256f850be6b1c8a21aec6e32dae16fd [https://perma.cc/2ABN-P6LC].

59. *See infra* Section I.B.1.

60. *See* JOSEPH TUROW, THE DAILY YOU: HOW THE NEW ADVERTISING INDUSTRY IS DEFINING YOUR IDENTITY AND YOUR WORTH 74 (2011); Dale Ganley & Cliff Lampe, *The Ties That Bind: Social Network Principles in Online Communities*, 47 DECISION SUPPORT SYS. 266, 266 (2009); Charles Kadushin, *The Friends and Supporters of Psychotherapy: On Social Circles in*

1. On Context and the Flow of Information

Multidisciplinary research addresses the three main factors that influence the information flow:[61] the source of the message, the context of the message, and the audience of the message.[62] Specifically, the likelihood that a message or idea will influence and spread depends on who posts the message and whether the source of the message is an influential hub in the social network; [63] the context of the message and the way it is presented; [64] and the audience and social structures of the message recipients in a given network. [65]

Intermediaries rely on existing research and also conduct experiments of their own.[66] They may even allow other stakeholders to conduct experiments on their platforms.[67] Understanding context and social graphs allows intermediaries to harness technologies and influence flows of information in many transparent and nontransparent ways.[68] For example, intermediaries can disseminate information through influential hubs in social networks and consequently influence

---

Urban Life, 31 AM. SOC. REV. 786, 801 (1966); Manoj Parameswaran & Andrew B. Whinston, *Research Issues in Social Computing*, 8 J. ASS'N FOR INFO. SYS. 336, 346 (2007); Kendra Cherry, *Skinner Box or Operant Conditioning Chamber*, VERY WELL MIND (Aug. 30, 2018), https://www.verywellmind.com/what-is-a-skinner-box-2795875 [https://perma.cc/W4M3-JBSR] (explaining the "skinner box"—an apparatus that can condition behavior); Whigham, *supra* note 58; discussion *infra* Section I.C.

61. *See* Michal Lavi, *Taking Out of Context*, 31 HARV. J.L. & TECH. 145, 150 (2017). Insights gleaned from psychology emphasize the power of context. *See id.* at 147. Findings suggest that context has more influence than the individuals who are engaged in an activity. *See* KAHNEMAN, *supra* note 48, at 171; THALER & SUNSTEIN, *supra* note 1, at 1–2; Philip G. Zimbardo, *The Journey from the Bronx to Stanford to Abu Ghraib*, *in* JOURNEYS IN SOCIAL PSYCHOLOGY: LOOKING BACK TO INSPIRE THE FUTURE 1, 34 (Robert Levine, Aroldo Rodrigues & Lynette Zelezny eds., 2008), http://pdf.prisonexp.org/PersonalJourney.pdf [https://perma.cc/2FGC-MX6L].

62. *See* Lavi, *supra* note 61, at 150.

63. *See* MALCOLM GLADWELL, THE TIPPING POINT: HOW LITTLE THINGS CAN MAKE A BIG DIFFERENCE 60 (2002) (referring to individuals who possess a great deal of information as "mavens"); CHARLES KADUSHIN, UNDERSTANDING SOCIAL NETWORKS: THEORIES, CONCEPTS, AND FINDINGS 144–46 (Deborah Grant ed., 2012) (referring to individual media influencers as "opinion leaders" and "influentials"); ELIHU KATZ & PAUL LAZARSFELD, PERSONAL INFLUENCE: THE PART PLAYED BY PEOPLE IN THE FLOW OF MASS COMMUNICATION 25 (1955); EVERETT M. ROGERS, DIFFUSION OF INNOVATION 27 (5th ed. 2003).

64. *See* GLADWELL, *supra* note 63, at 89; Jonah Berger & Katherine Milkman, *What Makes Online Content Viral?*, 49 J. MARKETING RES. 192, 192 (2012); Joseph E. Phelps et al., *Viral Marketing or Electronic Word-of-Mouth Advertising: Examining Consumer Responses and Motivations to Pass Along Email*, J. ADVERT. RES., Dec. 2004, at 333, 338.

65. *See* GLADWELL, *supra* note 63, at 158; KADUSHIN, *supra* note 63, at 146.

66. *See* BRETT FRISCHMANN & EVAN SELINGER, RE-ENGINEERING HUMANITY 117 (2018); James Grimmelmann, *The Law and Ethics of Experiments on Social Media Users*, 13 COLO. TECH. L.J. 219, 240, 263–64 (2015).

67. For example, Cambridge Analytica collected information on Facebook users and analyzed it, under the guise of academic research, to promote political purposes. *See* Samuel, *supra* note 19.

68. *See* SIVA VAIDHYANATHAN, ANTISOCIAL MEDIA: HOW FACEBOOK DISCONNECTS US AND UNDERMINES DEMOCRACY 58 (2018).

the message itself.[69] They can frame their platform in certain ways and change the context of messages that users publish.[70] By doing so, they also attract specific audiences and influence the nature of social ties among recipients.[71]

Intermediaries apply these insights to technology and sway users' decisions to generate and disseminate content.[72] Simple changes in the manner they design their platforms make a difference.[73] Small cues, or "channeling factors," can result in techno-social engineering and systematically lead individuals to change their behavior.[74]

Platform designs—the ways choices are presented to users, the number of choices, and the manner in which attributes are described—allow intermediaries to enhance cooperation, equality, and stability.[75]

---

69. *See* BRUCE SCHNEIER, DATA AND GOLIATH: THE HIDDEN BATTLES TO COLLECT YOUR DATA AND CONTROL YOUR WORLD 58 (2015); Sinan Aral & Dylan Walker, *Identifying Influential and Susceptible Members of Social Networks*, 337 SCI. 337, 337 (2012) (presenting a method that can identify profiles of "influential" members in given social networks). These insights are used in practice to promote campaigns and influence voters effectively. *See* COLIN BENNETT, VOTER SURVEILLANCE, MICRO-TARGETING AND DEMOCRATIC POLITICS: KNOWING HOW PEOPLE VOTE BEFORE THEY DO 3 (2014); VAIDHYANATHAN, *supra* note 68, at 172 (referring to Facebook's custom audiences service that allows advertisers efficient targeting); Daniel Kreiss, *Yes We Can (Profile You): A Brief Primer on Campaigns and Political Data*, 64 STAN. L. REV. ONLINE 70, 70 (2012); Jonathan Zittrain, *Engineering an Election*, 127 HARV. L. REV. F. 335, 335 (2014); Charles Duhigg, *Campaigns Mine Personal Lives to Get Out Vote*, N.Y. TIMES (Oct. 13, 2012), https://www.nytimes.com/2012/10/14/us/politics/campaigns-mine-personal-lives-to-get-out-vote.html [https://perma.cc/3C2E-WQPT]; *About Custom Audiences from Customer Lists*, FACEBOOK: BUS., https://www.facebook.com/business/help/341425252616329 [https://perma.cc/DQV7-X9XR] (last visited Sept. 26, 2018).

70. *See* HARTZOG, *supra* note 13, at 35. For example, messages on TheDirty.com are perceived to be negative because the name of the platform frames them as such. *See* Knibbs, *supra* note 9.

71. *See* Michal Lavi, *Content Providers' Secondary Liability: A Social Network Perspective*, 26 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 855, 856 (2016). On the importance of the strength of ties, see *id.*

72. *See* FRISCHMANN & SELINGER, *supra* note 66, at 18. Technology has an important role in influencing context. It creates affordances that can incline users to adopt different behaviors and pursue different paths of personal development. *See id.*; FOGG, *supra* note 15, at 5.

73. *See* Amos Tversky & Daniel Kahneman, *The Framing of Decisions and the Psychology of Choice*, 211 SCI. 453, 454 (1981). Individuals react to a particular choice in different ways depending on how it is presented. This is the "Framing Effect." *See* HARTZOG, *supra* note 13, at 35; KAHNEMAN, *supra* note 48, at 374; Tversky & Kahneman, *supra*, at 454.

74. *See* Howard Leventhal et al., *Effects of Fear and Specificity of Recommendation Upon Attitudes and Behavior*, 2 J. PERSONALITY & SOC. PSYCHOL. 20, 20–29 (1965). Channeling factors can change behavior. For example, when students are advised to get tetanus inoculations, they are far more likely to do so when given precise instructions where to go and what to do to get a shot. Giving students a map of the campus with the University Health Building circled and requesting to review their weekly schedule to locate a time to be inoculated made a difference. These small contextual changes channeled them towards a decision to get inoculations relative to other students who heard a lecture about the importance of inoculations. *See id.*

75. *See, e.g.*, Karen Levy & Solon Barocas, *Designing Against Discrimination in Online Markets*, 32 BERKELEY TECH. L.J. 1183, 1183, 1189, 1192 (2017) (providing a conceptual framework for understanding how platform design and policy choices introduce opportunities for users' biases to affect how they treat one another). The study focused on the influence of design on

However, choice architecture can also lead to antisocial behavior. Some business models are based on nudges that enhance extreme or offensive content to attract users and, in turn, increase advertising revenue.[76]

Nudges also affect the macrolevel of a social network. They can motivate sociological process and interpersonal dynamics by disseminating harmful speech across the network—enhancing the likelihood for widespread dissemination and causing tremendous harm to one's reputation.[77] This strategy may even incentivize the dissemination of fake news to promote commercial or political purposes,[78] manipulatively influence election results, or hinder political security in various ways.[79] The recent story of Facebook and Cambridge Analytica serves as a good example of the influences nudges can have on democracy.[80]

---

discrimination and analyzed ten categories of design and policy choices through which platforms may make themselves more or less conducive to discrimination by users. *See id.*

76. *See* JARON LANIER: TEN ARGUMENTS FOR DELETING YOUR MEDIA ACCOUNT RIGHT NOW 28 ,29 (2018) (coining the acronym BUMMER—Behaviors of Users Modified and Made into an Empire for Rent—to describe the influence of social media business models on users). *See also* VAIDHYANATHAN, *supra* note 68, at 5–6, 9 (describing how Facebook develops algorithms that favor highly charged or extremist content and depend upon a self-serving advertising system that precisely targets ads by using massive surveillance and elaborate personal dossiers). *But see* Peter Kafka, *YouTube is Trying to Clean Itself Up by Making It Much Harder for Small Video Makers to Make Money*, RECODE (Jan. 16, 2018, 6:00 PM), https://www.recode.net/2018/1/16/16898660/youtube-content-advertising-revenue-program-new-rules-google-preferred [https://perma.cc/W8ZY-S83F] (explaining how some YouTube advertisers want assurances that their content will be displayed next to brand-safe videos). On business models premised on generating offensive content, see DANIELLE KEATS CITRON, HATE CRIMES IN CYBERSPACE 6 (2014).

77. *See* KARINE NAHON & JEFF HEMSLEY, GOING VIRAL 20–21 (2013).

78. *See* SCHNEIER, *supra* note 69, at 54 (describing personalized microtargeting that can be used for commercial and electoral purposes).

79. *See* VAIDHYANATHAN, *supra* note 68, at 11 (explaining that "fake news," "propaganda," and "disinformation" result in the continual undermining of public trust in expertise, as well as rational deliberation and debate); Zittrain, *supra* note 69, at 335. One US election study involved Facebook users that were encouraged to click a button if they voted, which would create and share a post about their participation with a graphic sign and pictures of people who participated. Apparently, Facebook did not show these graphic signs to some users. Researchers cross-referenced names with actual voting records and found that those people who saw posts that their friends voted were more likely to vote. This study illustrates how intermediaries can influence voting rates and even election results. *See* Zittrain, *supra* note 69, at 335.

80. *See* VAIDHYANATHAN, *supra* note 68, at 55, 150; Carole Cadwalladr & Emma Graham-Harrison, *Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach*, GUARDIAN (Mar. 17, 2018, 6:03 PM), https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election [https://perma.cc/5JHM-4E8U].

C. Thresholds, Nudges, Sociological Process, and Dissemination of Speech

Decades ago, sociologists began to examine processes of information dissemination and developed models of social movements and collective behavior. In *Threshold Models of Collective Behavior*, Mark Granovetter articulated the key concept of a "threshold" to explain these processes.[81] A threshold is the number of people who must reach a single decision before a given actor follows.[82] The model assumes that information and ideas are considered more valuable as more individuals accept and adopt them.[83] A person's threshold for joining an activity is defined as "the proportion of the group a person would have to see join before he would do so."[84] In this way, a person's behavior depends upon the number of other people who already engage in that particular behavior. Therefore, one's social network has a huge potential to affect one's decisions to adopt and disseminate certain ideas.[85]

Each individual has a different threshold for adopting and disseminating ideas.[86] There are three categories of individual thresholds. First, "receptives" are individuals who have the lowest-level threshold for accepting new ideas.[87] Sometimes they already have a prior disposition in favor of a newly presented idea.[88] "Neutrals" have no inclination either way; yet, with a little information or exposure to a shared view of few people, they might come to accept an idea.[89] Finally, "skeptics" are individuals who have a high threshold for accepting ideas and might even hold a prior disposition standing in contrast to a newly presented idea.[90] These individuals require a great deal of information

81. Mark Granovetter, *Threshold Models of Collective Behavior*, 83 AM. J. SOC. 1420, 1422 (1978) (explaining that different individuals require different levels of safety for joining an activity, such as entering a riot, and also vary in the benefits they derive from the activity).

82. *See id.*

83. *See id.* at 1421.

84. *See id.* at 1422.

85. *See* NICHOLAS A. CHRISTAKIS & JAMES H. FOWLER, CONNECTED: THE SURPRISING POWER OF OUR SOCIAL NETWORKS AND HOW THEY SHAPE OUR LIVES 127 (2009); Lavi, *supra* note 71, at 889; Michal Shur-Ofry, *Popularity as a Factor in Copyright Law*, 59 U. TORONTO L.J. 525, 530 (2009).

86. *See* CASS SUNSTEIN, ON RUMORS: HOW FALSEHOODS SPREAD, WHY WE BELIEVE THEM, WHAT CAN BE DONE 19 (2009); Granovetter, *supra* note 81, at 1422.

87. SUNSTEIN, *supra* note 86, at 19 (explaining that the individual threshold depends on a person's prior disposition regarding the information).

88. *See id.*; Edward L. Glaeser & Cass R. Sunstein, *Does More Speech Correct Falsehoods?*, 43 J. LEGAL STUD. 65, 67 (2014) (explaining that because people have different prior beliefs, they will consequently have different degrees of skepticism).

89. *See* SUNSTEIN, *supra* note 86, at 20.

90. *See id.*

before accepting a given idea.[91] However, once the evidence becomes overwhelming—because the beliefs seem to be shared by many others—skeptics will join others in accepting the idea.[92]

Receptives are the first to adopt and circulate an idea. A subsequent reader then decides whether to adopt or decline the idea according to his personal threshold. The proliferation of an idea heavily depends on the types of individuals it encounters at the outset.[93] If an idea encounters an adequate number of receptives, neutral individuals are more likely to reach their threshold, and eventually, skeptics will finally follow and further spread the idea.[94]

When many individuals adopt an idea, positive feedback forms.[95] Thus, more individuals who are likely to reach their threshold follow them and further spread the idea. Diffusion of ideas, trends, or social behavior begins slowly.[96] When a critical mass of individuals publicly share an idea, a "tipping point" occurs, and the idea spreads like wildfire.[97]

Many times, an idea spreads to an "influential" individual in a social network. In such cases, if he accepts the idea and spreads it, the likelihood for reaching a tipping point exponentially increases.[98] The proliferation of an idea, thus, heavily depends on the individuals who encounter it at the outset.[99] An individual's threshold depends on various personal factors and social structures,[100] which may affect the collective outcome more than individual preferences.[101] The composition of a network as either homogenous or heterogeneous, thus, influences the extent of interdependence and bears on the likelihood of spreading an idea.[102]

---

91. *See id.*

92. *See id.*

93. *See id.* at 24.

94. *See id.* at 20.

95. *See id.*

96. *See id.*

97. *See* GLADWELL, *supra* note 63, at 12 (defining a tipping point as "the moment of critical mass, the threshold, the boiling point").

98. *See* KADUSHIN, *supra* note 63, at 146 (explaining the concept of "influentials").

99. *See id.* at 156.

100. *See id.* at 160–61 (referring to personal thresholds and exogenous factors, such as a network's structures and the proportion of adopters in one's direct interpersonal environment, as influences on people's decisions or actions).

101. *See* Granovetter, *supra* note 81, at 1430–31.

102. *See* Ronald S. Burt, *Social Contagion and Innovation: Cohesion versus Structural Equivalence*, 92 AM. J. SOC. 1287, 1290 (1987); Mark Granovetter & Roland Soong, *Threshold Models of Diffusion and Collective Behavior*, 9 J. MATHEMATICAL SOC. 165, 166 (1983) (focusing on the homogeneity assumption in models whereby the network is composed of homogenous individuals). In practice, people may vary from one another. Granovetter & Soong, *supra*, at 166.

Social networks are the key to understanding the flow and dissemination of information. Changes in an idea's composition, social structure, and transition path significantly alter the likelihood of widespread dissemination.[103] The results of psychological experiments demonstrate the influences social network structures have on the flow of information.[104] But sociological research reflects and explains these processes on the macrolevel of networks.[105] When a person with a low-level threshold adopts an idea, others are more likely to follow due to the network's influences.[106] This dynamic generates informational and reputational cascades that lead to an extensive dissemination of ideas throughout the network.[107] These insights frequently allow some prediction on human behavior and dynamics in a particular given network. The following Section explores prominent strategies of influences on social context and outlines an innovative taxonomy of nudge torts that can influence speech and incentivize widespread dissemination across the network.

D. Taxonomy of Online Evil Nudges

Granovetter wrote his seminal work over thirty years ago,[108] yet it can be smoothly applied to social media. Online networks operate in an environment designed by intermediaries.[109] Technological tools provide intermediaries with unilateral power to design architecture and

---

103. *See* KADUSHIN, *supra* note 63, at 157–61.

104. *See, e.g.*, Solomon E. Asch, *Studies of Independence and Conformity: A Minority of One Against a Unanimous Majority*, 70 PSYCHOL. MONOGRAPHS: GEN. & APPLIED 1, 1 (1956) (describing the "conformity experiment"—whereby study subjects provided wrong answers to questions to conform with the rest of the group). This phenomenon is defined as "herding." *See* CASS SUNSTEIN, GOING TO EXTREMES: HOW LIKE MINDS UNITE AND DIVIDE 57 (2009); Sushil Bikhchandani et al., *Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades*, 12 J. ECON. PERSP. 151, 152 (1998).

105. *See* DUNCAN J. WATTS, SIX DEGREES: THE SCIENCE OF A CONNECTED AGE 208 (2003); Lev Muchnik et al., *Social Influence Bias: A Randomized Experiment*, 341 SCI. 647, 647–51 (2013) (reporting experiment results where participants preferred to download files that were already downloaded by others); Muchnik et al., *supra*, at 647 (explaining that positive or negative ratings in review websites influence participants); Matthew Salganik et al., *Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market*, 311 SCI. 854, 854 (2006).

106. *See* Granovetter, *supra* note 81, at 1422.

107. Informational cascades are generated when individuals follow the statements or actions of predecessors and do not express their opposing opinions because they believe their predecessors are right. As a result, the social network does not obtain important information. Reputational cascades form because of social pressures. In these cases, people think they know what is right, or what is likely to be right, but they nonetheless go along with the crowd in order to maintain their status. *See* SUNSTEIN, *supra* note 104, at 57; Cass R. Sunstein & Reid Hastie, *Four Failures of Deliberating Groups* 2 (Univ. of Chi. Pub. Law, Working Paper No. 215, 2008).

108. *See* Granovetter, *supra* note 81, at 1420.

109. *See* Ryan Calo, *Digital Market Manipulation*, 82 GEO. WASH. L. REV. 995, 995 (2014).

influence decision-making from the top-down.[110] Intermediaries influence context in a variety of ways, using strategies in accordance with a myriad of business models.[111] Unlike typical models of traditional media, these models of influence do not only involve a passive audience, but rather they affect active users who disseminate the information as well.[112] Furthermore, data collection,[113] complex algorithms,[114] and technologies—such as machine learning, artificial intelligence (AI),[115] big data,[116] and IoT[117]—allow intermediaries to use data, hack the human consciousness, and enhance their influence.[118] Intermediaries can identify influential hubs in a given network[119] and target particular users.[120] They can conduct complex studies on user

110. *See* HARTZOG, *supra* note 13, at 34 (explaining that design can shape our perceptions, behavior and values).

111. *See id.* (referring to intermediaries' power to design architecture and personalize information); Calo, *supra* note 109, at 995. For more on personalization and optimization of relationships, see FRISCHMANN & SELINGER, *supra* note 66, at 150.

112. *See* Calo, *supra* note 109, at 1042.

113. *See* VAIDHYANATHAN, *supra* note 68, at 55 (giving the example of Facebook, which has grown into the "most pervasive surveillance system in the world" and also the "most reckless and irresponsible surveillance system in the commercial world").

114. *See id.* at 150 (referring to algorithmic targeting in the 2016 elections and raising the question of what democracy would look like if algorithms governed the art of science and persuasion); *id.* at 150–55 (referring to psychographic profiling that allows accurate algorithmic targeting).

115. *See* Balkin, *supra* note 54, at 1184; Karen Yeung, *'Hypernudge': Big Data as a Mode of Regulation by Design*, INFO. COMM. & SOC'Y, May 2016, at 8 (nudging can be integrated into machine learning techniques); Hayley Tsukayama, *Facebook is Using AI to Try to Prevent Suicide*, WASH. POST (Nov. 27, 2017), https://www.washingtonpost.com/news/the-switch/wp/2017/11/27/facebook-is-using-ai-to-try-to-prevent-suicide/?utm_term=.93d758864ed9 [https://perma.cc/9YAC-F5MA]. For expansion on AI in general, see ADAM THIERER ET AL., MERCATUS CTR., GEORGE MASON UNIV., ARTIFICIAL INTELLIGENCE AND PUBLIC POLICY 2 (2017); Ryan Calo, *Artificial Intelligence Policy: A Primer and Roadmap*, 51 U.C. DAVIS L. REV. 399, 404 (2017). For negative usage of AI to spread harmful speech, see FUTURE OF HUMANITY INST., *supra* note 20, at 1, 3; Meg Leta Jones, *Silencing Bad Bots: Global, Legal and Political Questions for Mean Machine Communication*, 23 COMM. L. & POL'Y 159, 185 (2018) (addressing the negative usage of AI to spread harmful speech by bots).

116. *See* Hannes Grassegger & Mikael Krogerus, *The Data That Turned the World Upside Down*, MOTHERBOARD (Jan. 28, 2017, 8:15 AM), https://motherboard.vice.com/en_us/article/mg9vvn/how-our-likes-helped-trump-win [https://perma.cc/CJP4-TR89].

117. Today, sensors in physical objects can passively collect information on individuals and their networks online and offline. *See* MIREILLE HILDEBRANDT, SMART TECHNOLOGIES AND THE END(S) OF LAW: NOVEL ENTANGLEMENTS OF LAW AND TECHNOLOGY 9, 41 (2015); VAIDHYANATHAN, *supra* note 68, at 101 (explaining that specific technologies and intermediaries interact with users' minds and bodies).

118. YUVAL NOACH HARRARI, 21 LESSONS FOR THE 21ST CENTURY 267–68 (2018) ("You might have heard that we are living in the era of hacking computers, but that's hardly half the truth. In fact, we live in the era of hacking humans.").

119. On "influentials," see KADUSHIN, *supra* note 63, at 146. *See also* Aral & Walker, *supra* note 69, at 337 (finding an innovative way to measure influence in decisions to adopt products in a given social network).

120. *See* Aral & Walker, *supra* note 69, at 340.

behavior[121] and nudge more efficiently by harnessing multidisciplinary insights and technological tools that make it easier to influence behavior online in a persuasive way.[122] Nudges increase the likelihood that individuals reach their threshold to support an idea, repeat the idea, and spread it further.[123]

As an idea circulates, it tends to gain credibility. The more people repeatedly hear an idea, the more likely they are to believe that idea.[124] Nudging users to disseminate defamation and false information reduces the likelihood for a successful correction of erroneous information by other members of the social network.[125] Repeating informational errors not only exacerbates harm, but also undermines efficient bottom-up private ordering by participants that commonly outline and enforce social norms.[126] Moreover, researchers have revealed that falsehoods are disseminated significantly farther, faster, deeper, and more broadly than the truth.[127]

Due to the potential harm that an intermediary's influence may inflict, comprehensive theoretical analysis of the liability of intermediaries in speech tort is indispensable. The following

---

121. *See* FRISCHMANN & SELINGER, *supra* note 66, at 117 (describing the cognition experiment of Facebook on users' emotions); VAIDHYANATHAN, *supra* note 68, at 154–55 (referring to psychometrics data-driven personality quizzes used by Cambridge Analytica); Tsukayama, *supra* note 115 (noting that Facebook may use AI to understand user emotions and identify situations requiring intervention); Whigham, *supra* note 58 (describing a Facebook experiment concerning the influencing of susceptible minors).

122. *See* THALER, *supra* note 32, at 341–42. On intermediaries and other stakeholders influences on consumers by using IoT technologies, see JOSEPH TUROW, THE AISLES HAVE EYES: HOW RETAILERS TRACK YOUR SHOPPING, STRIP YOUR PRIVACY, AND DEFINE YOUR POWER 18–19 (2017). It should be noted that IoT technologies may also similarly influence speech. *See* FRISCHMANN & SELINGER, *supra* note 66, at 11 ("[I]t's rapidly becoming easier to design technologies that nudge us to go on auto pilot and accept the cheap pleasure that comes from minimal thinking. Smart environments are poised to significantly exacerbate this situation."); HARTZOG, *supra* note 13, at 146 ("Companies have learned that targeted, personalized appeals are more persuasive than ads designed for a general audience.").

123. *See* THALER & SUNSTEIN, *supra* note 1, at 65–66 (describing a cascade triggered by nudges).

124. *See* NICHOLAS DIFONZO & PRASHANT BORDIA, RUMOR PSYCHOLOGY: SOCIAL AND ORGANIZATIONAL APPROACHES 225 (2007); SUNSTEIN, *supra* note 86, at 21; Gordon Pennycook et al., *Prior Exposure Increases Perceived Accuracy of Fake News*, J. EXPERIMENTAL PSYCHOL. (forthcoming) (manuscript at 4). The recent campaign elections in the United States illustrate this proposition. For example, it was rumored that Pope Francis endorsed Donald Trump, even though he has done nothing of the sort. Some commentators claim that repeated exposure to this falsehood and others like it influenced the results. *See* ZEYNEP TUFEKCI, TWITTER AND TEAR GAS: THE POWER AND FRAGILITY OF NETWORKED PROTEST 264–65 (2017); Zeynep Tufekci, *Mark Zuckerberg Is in Denial*, N.Y. TIMES (Nov. 15, 2016), http://www.nytimes.com/2016/11/15/opinion/mark-zuckerberg-is-in-denial.html [https://perma.cc/RXN9-VB29].

125. *See* Bikhchandani et al., *supra* note 104, at 165, 168.

126. *See* Lavi, *supra* note 71, at 914, 916. For expansion on private ordering of speech related harm online, see generally *id.*

127. *See* Soroush Vosoughi, Deb Roy & Sinan Aral, *The Spread of True and False News Online*, 359 SCIENCE 1146 (2018) (finding that false tweets spread faster than true tweets).

Subsections review several types of intermediary nudges, demonstrate their harm potential, and classify nudges into the following categories: (1) "focal point," (2) "channeling and leading," and (3) "encouragement." These categories form a descriptive taxonomy for understanding nudges in social networks. All of these strategies lead to similar results, yet each has nuances and distinctive characteristics.

Focal point nudges influence a potential user's decision to participate in an online platform, whereas channeling and leading and encouragement nudges influence users only after the decision to participate has been made.[128] Moreover, the extent of transparency with respect to the goals of the intermediary is not uniform. Some nudges appeal to conscious, deliberative thinking, while others influence in nonsalient, or nontransparent ways—leading to nondeliberative decision-making, as well as subconsciously influencing information processing.[129] The taxonomy constitutes a significant contribution by identifying these strategies and their effects on speech torts, fake stories, and other types of harmful speech.

This taxonomy focuses on the main types of nudges and does not purport to encompass all the possible tactics of influence. Indeed, more strategies may develop as technologies advance. Yet, by mapping the main nudges and understanding their effects on social networks, updating future changes should become easier.

1. Focal Point

- *A popular website is titled The Dirty.*[130]
- *Famous rating and review websites are titled Ripoff Report, BadBusiness, and PissedConsumer.*[131]

Various research fields have defined the idea of focal point or foci in a number of ways.[132] In a broader context, researchers define it as a market around which people can organize their affairs without explicit communication.[133] Sociologists of networks define foci as "social, psychological, or physical entit[ies] around which joint activities are organized ([such as] workplaces, voluntary organizations, . . . families,

128. See LIOR JACOB STRAHILEVITZ, INFORMATION AND EXCLUSION 42–45 (2011). Focal points frame the platform itself and therefore can affect the very decision to participate, whereas channeling and leading and encouragements are elements inside the platform; therefore, they have no influence on the decision to participate and the composition of the participants.

129. *See* Sunstein, *supra* note 24 (manuscript at 36).

130. *See* THE DIRTY, *supra* note 9.

131. Courts have discussed the liability of these websites. *See infra* Section III.A.

132. *See* STRAHILEVITZ, *supra* note 128, at 42–45.

133. *See id.*

etc.)."[134] Individuals who organize themselves around the same foci tend to form a dense cluster of interpersonal ties.[135] These structures influence the context of social networks.[136]

The name of a website, its design, and explicit or implicit goals may constitute a focal point and affect the flow of information.[137] First, a focal point influences the composure and character of the individuals that choose to take part in conversations.[138] It brings participants with similar dispositions and preferences into one virtual space,[139] and conversely sends exclusionary signals to those participants with differing views.[140] Thus, it creates an "inclusionary vibe" among some individuals and, at the same time, an "exclusionary vibe" among others.[141] Second, the focal point affects the content that participants create and share.[142] It facilitates particular flows of information and forms a market for certain type of ideas.[143]

An intermediary that forms a focal point signals similarly situated individuals with comparable prior dispositions to participate, and pushes them to generate specific types of content.[144] Encountering resistance from other participants is less likely in a homogeneous network.[145] Consequently, the likelihood to cross individual and collective thresholds for spreading content increases.[146] This context

---

134. *See* Scott L. Feld, *The Focused Organization of Social Ties*, 86 AM. J. SOC. 1015, 1016 (1981).

135. *See id.*

136. *See id.* at 1033; Kadushin, *supra* note 60, at 786.

137. *See* Judith Donath, *Signals in Social Supernets*, 13 J. COMPUTER-MEDIATED COMM. 231, 235 (2008). This phenomenon is known in literature as a "halo effect." *See The Halo Effect*, ECONOMIST (Oct. 14, 2009), https://www.economist.com/news/2009/10/14/the-halo-effect [https://perma.cc/6H7V-R8UV].

138. *See* Donath, *supra* note 137, at 235.

139. *See* STRAHILEVITZ, *supra* note 128, at 44.; SUNSTEIN, *supra* note 104, at 75–79; CASS R. SUNSTEIN, #REPUBLIC: DIVIDED DEMOCRACY IN THE AGE OF SOCIAL MEDIA 123 (2017) (explaining that a prior disposition influences the likelihood of crossing the threshold); Glaeser & Sunstein, *supra* note 88, at 66, 91.

140. *See* STRAHILEVITZ, *supra* note 128, at 43; Lavi, *supra* note 71, at 928–29.

141. *See* STRAHILEVITZ, *supra* note 128, at 44.

142. *See* THALER & SUNSTEIN, *supra* note 1, at 36. The influence of Focal Points on the content created is a consequence of the framing effect created by the specific naming of the platform. *See id.* (expanding on the "framing effect," whereby choices depend, in part, on the way problems are stated).

143. *See id.* at 37. The focal point's exclusionary vibe and framing effect shape the market of ideas in the specific platform. STRAHILEVITZ, *supra* note 128, at 44; THALER & SUNSTEIN, *supra* note 1, at 37.

144. *See* Donath, *supra* note 137, at 237.

145. *See* Lavi, *supra* note 71, at 928–29.

146. *See* SUNSTEIN, *supra* note 104, at 80; Noah P. Mark, *Culture and Competition: Homophily and Distancing Explanation for Cultural Niches*, 68 AM. SOC. REV. 319, 335 (2003). Individual behavior is extreme and polarized when clustered with like individuals. *See* SUNSTEIN, *supra* note 139, at 123–24, 236, 243 (expanding on homogeneity and incitement for violence and terror).

engineers social norms[147] and enhances dissemination of particular types of ideas.[148] Focal point nudges are unique since they influence the decision to participate in a conversation on a given online platform.[149] This shapes the composition of participants, and the influence on content is only a byproduct.

a. Nuances of Focal Points and Gravity of Harm

Focal points take multiple forms and shades. Intermediaries can harm third parties by framing their platform in a manner that explicitly invites harmful content; for example, a platform entitled TheDirty.com[150] or HarrassThem.com are exemplars of such framing.[151] The business model employed by these intermediaries nudges tortious defamatory speech.[152]

Likewise, gossip website intermediaries signal to participants to do just that: gossip. For example, the now-defunct platform called JuicyCampus[153] pushed homogenous participants to spread rumors. This platform signaled that gossip was legitimate—bringing participants with similar preferences into the platform and removing social constraints.[154] As a result, the rate of false rumors and fake

147. *See* Michiru Nagatsu, *Social Nudges: Their Mechanisms and Justification*, 6 REV. PHIL. & PSYCHOL. 481, 488–89 (2015).

148. *See* KADUSHIN, *supra* note 63, at 158; Granovetter, *supra* note 81, at 1423. The focal point generates an exclusionary vibe and leads to clustering of individuals with low thresholds for accepting and diffusing specific types of content. Thus, the focal point increases the likelihood for the first adopter to cross the threshold and drive a sociological process of mass diffusion. *See* STRAHILEVITZ, *supra* note 128, at 44; SUNSTEIN, *supra* note 139, at 123–24 (explaining the phenomenon of "confirmation bias"—whereby similar people tend to confirm each other's speech in the process of diffusion).

149. The influence of the composition of participant is due to the exclusionary vibe that results from the focal point. On exclusionary vibes, see STRAHILEVITZ, *supra* note 128, at 44.

150. *See* THE DIRTY, *supra* note 9.

151. *See* Fair Hous. Council v. Roommates.com, LLC, 489 F.3d 921, 928 (9th Cir. 2007) (describing a hypothetical website for nudging harassment: "Imagine, for example, www.Harrassthem.com with the slogan "Don't Get Mad, Get Even." A visitor to this website would be encouraged to provide private, sensitive and/or defamatory information about other[s] . . . . In addition, the website would encourage the poster to provide dirt on the victim, with instructions that the information need not be confirmed but could be based on rumors, conjecture or fabrication").

152. *See* LORI ANDREWS, I KNOW WHO YOU ARE AND I SAW WHAT YOU DID: SOCIAL NETWORKS AND THE DEATH OF PRIVACY 105–09 (2012); Skyler McDonald, Note, *Defamation in the Internet Age: Why Roommates.com Isn't Enough to Change the Rules for Anonymous Gossip Websites*, 62 FLA. L. REV. 259, 271 (2010).

153. Matt Ivester, a Duke University alumnus, founded JuicyCampus. The site encouraged users to "Keep It Juicy" and vote on the "juiciest" posts. *See* Ali Grace Zieglowsky, *Immoral Immunity: Using a Totality of the Circumstances Approach to Narrow the Scope of Section 230 of the Communications Decency Act*, 61 HASTINGS L.J. 1307, 1320 (2010).

154. *See id.*

stories had the potential to increase.[155] Yet the focal point for gossip does not explicitly nudge participants to spread falsehoods; innocent gossip may also be welcome. These exchanges of personal information may have benefits.[156] However, a platform entitled JuicyCampus is likely to contain more inaccuracies, defamation, and falsehoods than a platform titled "Students" because requesting users to post "juicy" stories is likely to cause users to present the information in an inaccurate way. As a result, even a true story may be taken out of context and turned into defamation.[157]

Intermediaries can also nudge negative speech. Platforms designed for consumer complaints—such as BadBusiness.com, RipoffReport.com, or PissedConsumer.com[158]—push individuals to post negative reviews. These websites are focal points for negative reviews; however, they do not specifically push participants to spread defamatory content.[159] These platforms can improve the marketplace and prevent consumers' engagement in inefficient transactions.[160] However, they only draw unsatisfied consumers and implicitly exclude satisfied ones.[161] This homogenous social composition increases the likelihood for defamatory speech relative to websites that are "neutral" to negative content.[162]

Focal points unite homogeneous individuals with similar prior dispositions and push them to generate specific types of content.[163] Even an explicit nudge that invites participants to gossip or generate

---

155. *See* Brian McNeill, *Uva Student Council Unhappy with JuicyCampus.com*, DAILY PROGRESS (Mar. 26, 2008), http://www.dailyprogress.com/archives/uva-student-council-unhappy-with-juicycampus-com/article_d1ed4c24-61c3-5143-afe5-b6680f6c0a49.html [https://perma.cc/M63Z-NZN8].

156. *See* Diane L. Zimmerman, *Requiem for a Heavyweight: A Farewell to Warren and Brandeis's Privacy Tort*, 68 CORNELL L. REV. 291, 334 (1983) ("By providing people with a way to learn about social groups to which they do not belong, gossip increases intimacy and a sense of community among disparate individuals and groups.").

157. *See* Lavi, *supra* note 61, at 156.

158. *See* GW Equity LLC v. Xcentric Ventures LLC, No. 3:07-CV-976-O, 2009 WL 62173, at *3 (N.D. Tex. Jan. 9, 2009); Hy Cite Corp. v. Badbusinessbureau.com, LLC, 418 F. Supp. 2d 1142, 1149 (D. Ariz. 2005); Vo Grp., LLC v. Opinion, Corp., No. 8758/11, at 11–12 (N.Y. App. Div. May 22, 2012); discussion *infra* Section III.A.1.

159. *See* Zieglowsky, *supra* note 153, at 1326.

160. *See* Eric Goldman, *Expert Report on the Value of Consumer Review Websites and 47 USC 230*, TECH. & MARKETING L. BLOG (Nov. 20, 2012), http://blog.ericgoldman.org/archives/2012/11/expert_report_o.htm [https://perma.cc/33SM-3MZZ] (explaining that the mechanism of punishing bad producers depends on well-informed consumers).

161. *See* Kristine L. Gallardo, Note, *Taming the Internet Pitchfork Mob: Online Public Shaming, the Viral Media Age, and the Communications Decency Act*, 19 VAND. J. ENT. & TECH. L. 721, 723–24 (2017); Goldman, *supra* note 160.

162. *See* Gallardo, *supra* note 161, at 723. The starting point in this platform is composition of people who intend to publish negative information as opposed to neutral starting point. *See* Feld, *supra* note 133, at 1016.

163. *See* Feld, *supra* note 134, at 1016.

negative content may be broadly interpreted—thus exacerbating the severity of harm.[164] The homogenous composition of users increases the likelihood that participants cross the threshold for adopting and sharing tortious, gossipy, and negative content.[165] When a critical mass of people adopt and disseminate this type of content, a tipping point is created, and the content takes off and spreads like wildfire.[166] In addition to increasing the proportion of negative speech, this sociological dynamic leads to polarization and extremism, and enhances the strength and influence of the offensive content.[167]

This dynamic undermines the likelihood for private ordering.[168] First, the social context decreases the potential to counter falsehoods by the victims.[169] Second, the homogenous composition of participants increases the likelihood that they mutually validate the content.[170] As a result, speech-related harm is exacerbated. Therefore, the gravity of the harm depends largely on the degree of the focal point of a nudge.

2. Channeling and Leading

- *An online intermediary of a review website for rating hotels requires participants to choose between two options in the platform's menu as a title to their review: "reasonable hotel" or "awful hotel."*[171]
- *An online intermediary of a review website requires users to categorize their reviews. Most of the categories offered are*

---

164. *See* Lavi, *supra* note 61, at 156.

165. *See* SUNSTEIN, *supra* note 104, at 75; SUNSTEIN, *supra* note 139, at 123; Glaeser & Sunstein, *supra* note 88, at 66, 91. The exclusionary vibe created by the focal point increases confirmation bias among participants that enforce their beliefs and influences the content. *See* STRAHILEVITZ, *supra* note 128, at 44; SUNSTEIN, *supra* note 139, at 123.

166. On informational and reputational cascades, see GLADWELL, *supra* note 63, at 12; KADUSHIN, *supra* note 63, at 136–37; CASS R. SUNSTEIN, INFOTOPIA: HOW MANY MINDS PRODUCE KNOWLEDGE 92 (2006); SUNSTEIN, *supra* note 104, at 79.

167. *See* Marcial Losada & Emily Heaphy, *The Role of Positivity and Connectivity in the Performance of Business Teams: A Nonlinear Dynamics Model*, 47 AM. BEHAV. SCIENTIST 740, 761 (2004). The quantity of defamatory speech may affect the strength of each expression. *See id.*

168. *See* Lavi, *supra* note 71, at 928–29 (describing challenges to private ordering in networks of homogenous participants). The context of the network created within negative focal points is not neutral because the starting point is a composition of users that have negative prior beliefs on business. Consequently, there is even less likelihood for private ordering.

169. *See* N.Y. Times Co. v. Sullivan, 376 U.S. 254, 270 (1964); Robert D. Richards & Clay Calvert, *Counterspeech 2000: A New Look at the Old Remedy for "Bad" Speech*, 2000 BYU L. REV. 553, 555 (2000) (explaining that in some cases, speech can be countered).

170. *See* Cohen, *supra* note 57, at 150 (explaining that homogenous groups can more easily become polarized in their beliefs and perceptions of reality).

171. In this situation, the intermediary provides *only* unlawful titles. It can be argued that this situation is similar to the case of *Fair Hous. Council v. Roommates.com, LLC*, 489 F.3d 921, 924, 926 (9th Cir. 2007) (finding that the operator of the website forced user to provide discriminatory answers to drop-down menu questions as a pre-condition for participation).

negative, such as "rip off," "con artists," and "corrupt companies."[172]

- *An online intermediary of a review website installs filters on its website. These filters allow users to view specific parts of the information on the platform. The default option allows only negative reviews and filters out the rest.*[173]

Herbert Simon pointed out that a "wealth of information creates a poverty of attention."[174] This statement is of great relevance today.[175] One of the most important challenges of intermediaries in the digital era is to direct user attention and assist with focusing on the most relevant content.[176] To meet this challenge, intermediaries design mechanisms to assist users in navigating the growing sea of information.[177] At times, intermediaries use the cognitive biases of users and channel them to support consumption and dissemination of specific types of content to enhance their profits.[178] Drop-down menus, default rules, tagging options, and filtering mechanisms can channel and lead to particular choices.[179]

A high rate of negative or defamatory options creates a "framing effect"[180] and a "priming effect,"[181] which increase the likelihood that a user will choose one of the negative or defamatory options.[182] This may increase the generation and dissemination of tortious and negative content.[183] Nudging choices through default options, which allow users to see only negative reviews and filter out the rest, usually leads to a

---

172. *See* Glob. Royalties, Ltd. v. Xcentric Ventures, LLC, 544 F. Supp. 2d 929, 930 (D. Ariz. 2008).

173. For a similar case, see *Fair Hous. Council*, 489 F.3d at 929.

174. Herbert A. Simon et al., *Designing Organizations for an Information-Rich World, in* COMPUTERS, COMMUNICATIONS, AND THE PUBLIC INTEREST 37, 40–41 (M. Greenberger ed., 1971).

175. *See* HOWARD RHEINGOLD, NET SMART: HOW TO THRIVE ONLINE 36 (2012).

176. *See id.* at 77; VAIDHYANATHAN, *supra* note 68, at 80 (referring to the attention economy and the important function of managing and filtering information); Kreimer, *supra* note 56, at 17.

177. *See* Stuart W. Shulman, *The Internet Still Might (But Probably Won't) Change Everything*, 1 I/S: J.L. & POL'Y FOR INFO. SOC'Y 111, 118 (2005) (explaining that user-friendly technological and procedural innovations, such as automatic categorizing, mitigate the problem of information overload).

178. *See* GREENFIELD, *supra* note 23, at 139 (explaining that information and design advantages translate into systematic consumer vulnerability in digital markets); Calo, *supra* note 109, at 999.

179. *See* HARTZOG, *supra* note 13, at 26, 161–62 (explaining how design can result in a myopia regarding items that are not included in the design of menus); *see also id.* (giving an example of a design that channels users to consent to terms that were not necessarily received under different design).

180. *See* Nagatsu, *supra* note 147, at 489–93.

181. *See id.*; KAHNEMAN, *supra* note 48, at 119–24; *infra* note 188 and accompanying text.

182. *See* KAHNEMAN, *supra* note 48, at 56, 58.

183. *See id.*

"status quo bias." As a result, users are not likely to deviate from them.[184]

Similar to focal points, channeling and leading nudges operate on both micro and macro levels. They enhance negative, defamatory content and increase the likelihood of extensive adoption and dissemination of content within the social network.[185] Nevertheless, unlike focal points, channeling and leading nudges do not influence the decision to use the platform.[186] This category of nudges is less transparent, and the intention of the intermediary to encourage specific types of content is less obvious.

a. Nuances of Channeling and Leading and Gravity of Harm

Intermediaries channel and lead users to distribute defamatory or negative content in various ways.[187] Intermediaries may include limited options—for example, providing two extreme options without offering a third—which can prime tortious or negative content,[188] and influence users' content.[189] Biased intermediary options that tilt closer to the negative side of the scale can also increase a user's likelihood to distribute negative content.[190] Intermediaries can also frame specific choices by focusing on their advantages and reminding participants of what they turn down by opting for the nonpreferred alternative.[191] In the same manner, internal search engines and filtering mechanisms can channel and lead users to generate, consume, and disseminate specific types of content.[192] Intermediaries can design dynamic menus

184. *See* HARTZOG, *supra* note 13, at 39 ("[F]rames that comport with the existing schemata in a receiver's belief system can be particularly effective."); Daniel Kahneman et al., *Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias*, 5 J. ECON. PERSP. 193, 197–98 (1991) (explaining that individuals have a strong tendency to remain at the status quo, because the disadvantages of leaving it loom larger than the advantages).

185. *See* KADUSHIN, *supra* note 63, at 137.

186. *See id.* at 139–42.

187. On choice architecture and default rules, see CASS R. SUNSTEIN, IMPERSONAL DEFAULT RULES VS. ACTIVE CHOICES VS. PERSONALIZED DEFAULT RULES: A TRIPTYCH 1, 3, 9 (2013).

188. For example, in *Fair Hous. Council v. Roommates.com*, the court decided to impose liability on an intermediary for designing a drop-down menu that channeled users to generate discriminatory content. *See* Fair Hous. Council v. Roommates.com, LLC, 521 F.3d 1157, 1169 (9th Cir. 2008).

189. *See* KAHNEMAN, *supra* note 48, at 119–24 (explaining that anchoring suggestions, as in the case of menus and tagging options, result in a "priming effect").

190. Biased intermediary options are likely to create an anchoring effect with regard to the middle option. Information processing that starts with a biased anchor is likely to lead users to adjust in that direction. For expansion on anchoring, see THALER & SUNSTEIN, *supra* note 1, at 23.

191. *See* Punam Anand Keller et al., *Enhanced Active Choice: A New Method to Motivate Behavior Change*, 21 J. CONSUMER PSYCHOL. 376, 378 (2011).

192. Intermediaries of review websites can channel users to consume only popular or negative reviews.

in which the selections are dependent upon some other input, such as a user's selection in a prior list, thus resulting in an even narrower scale of options.[193] Furthermore, technological tools allow intermediaries to receive and analyze users' personal information.[194] By using data mining, big data, AI, and personalizing default rules,[195] intermediaries exhibit different choices to different users and channel them more efficiently.[196]

Channeling and leading nudges influence the severity of harm to varying degrees. They facilitate the generation of content that the intermediary prefers on its platform.[197] For example, an individual user that intends to write a negative review may be primed to choose an extreme title such as "rip off" or "con artists."[198] This framing is likely to influence a user to write a more extreme review than he had first intended. By enhancing the distribution of negative or tortious content, the intermediary increases the likelihood that more users will cross the threshold for adopting and disseminating such content.

This strategy of nudges may also undermine the likelihood for private ordering. In contrast to focal points, channeling and leading nudges do not affect the composition of users; the population of users can be either homogenous or heterogeneous.[199] Nevertheless, the magnitude of negative content undermines the likelihood for correction by the victim or other users.[200]

3. Encouragement

- *An online intermediary encourages users to publish rumors, gossip, and defame by using slogans, such as "Don't let them get away with it! Let the truth be known!" It also states*

193. On dynamic drop-down menus, see Hattie Harman, *Drop-Down Lists and the Communications Decency Act: A Creation Conundrum*, 43 IND. L. REV. 143, 172 (2009).

194. *See id.* at 150–51, 172.

195. *See* Ariel Porat & Lior Strahilevitz, *Personalizing Default Rules and Disclosure with Big Data*, 112 MICH. L. REV. 1417, 1417 (2014); Shoshana Zuboff, *Big Other: Surveillance Capitalism and the Prospects of an Information Civilization*, 30 J. INFO. TECH. 75, 75 (2015) ("[Surveillance capitalism] aims to predict and modify human behavior as a means to produce revenue and market control.").

196. *See* ELI PARISER, THE FILTER BUBBLE: WHAT THE INTERNET IS HIDING FROM YOU 70–83 (2011); TUROW, *supra* note 60, at 110.

197. *See* TUROW, *supra* note 60, at 110.

198. *See* Glob. Royalties, Ltd. v. Xcentric Ventures, LLC, 544 F. Supp. 2d 929, 930 (D. Ariz. 2008).

199. *See* Richard H. Thaler, *Behavioral Economics: Past, Present, & Future*, 106 AM. ECON. REV. 1577, 1593–94 (2016). Channeling and leading nudges are elements within the platform and have no influence on the decision to participate in the first place. *See* Harman, *supra* note 193, at 172.

200. *See generally* Muchnik et al., *supra* note 105.

"Complaints[,] Reviews Scams[,] Lawsuits[, and] Frauds Reported."[201]

- *An online intermediary encourages users to publish rumors, gossip, and defamatory statements by using the slogan "Keep It Juicy."*[202] *The encouragement is applied generally and personally.*[203]
- *An online intermediary uses the slogan "Pure Evil" and encourages users to ruin the reputation of third parties.*[204] *Similarly, the owner of the website The Dirty encourages readers to email him "dirt" on people they know.*[205]
- *An online intermediary harvests profiles from Facebook and encourages users to make negative comments about social media users.*[206]

Intermediaries may explicitly or implicitly signal to users that specific types of content are desired on their platform, push users to publish that type of content, and intensify their distribution.[207] They use various strategies of encouragement to increase their influence.

---

201. *See* Vision Sec., LLC v. Xcentric Ventures, LLC, No. 2:13-CV-00926, 2015 WL 12780892, at *2 (D. Utah Aug. 27, 2015); RIPOFF REPORT, https://www.ripoffreport.com [https://perma.cc/38TX-LTGG] (last visited Sept. 29, 2018).

202. This was the slogan of the intermediary JuicyCampus.com. This intermediary generated a focal point nudge from its name and encouragement nudge by the slogan. *See* McNeill, *supra* note 155. In this specific platform, the encouragement was general and applied equally to all participants. *See id.* Yet, encouragements can be personalized.

203. *See* FRISCHMANN & SELINGER, *supra* note 66, at 150 ("If you bought books or music on Amazon . . . or even typed a message, you've engaged with machines that are designed to figure out how our minds work and steer our choices with personalized recommendations."). For personalized suggestions using big data and AI in a related context, see *Dyroff v. Ultimate Software Group, Inc.*, No. 17-cv-05359-LB, 2017 WL 5665670, at *2 (N.D. Cal. Nov. 26, 2017).

204. This was the slogan of the revenge porn platform IsAnyoneUp.com, which Hunter Moore founded. *See* Emily Poole, Comment, *Back Against Non-Consensual Pornography*, 49 U.S.F. L. REV. 181, 182 (2014).

205. *See* Caitlin Dewey, *The Government Just Took a Huge Step in the Fight Against Revenge Porn*, WASH. POST (Jan. 30, 2015), https://wapo.st/1BzHY7h?tid=ss_mail&utm_term=.16fce95c79f7 [https://perma.cc/DMF3-LMFR]; Kashmir Hill, *The Dirty Business: How Gossipmonger Nik Richie of TheDirty.com Stays Afloat*, FORBES (Nov. 11, 2010), https://www.forbes.com/sites/kashmirhill/2010/11/11/the-dirty-business-how-gossipmonger-nik-richie-of-thedirty-com-stays-afloat/#31afa6962f9b [https://perma.cc/XF9B-33QF].

206. *See* Fanning v. Fed. Trade Comm'n, 821 F.3d 164, 169 (1st Cir. 2016). The intermediary "jurk.com" took information from Facebook and encouraged users to label millions a "Jerk" or "not a Jerk." *Id.*

207. *See* Jeffrey R. Doty, *Inducement or Solicitation? Competing Interpretations of the "Underlying Illegality" Test in the Wake of Roommates.com*, 6 WASH. L. J. TECH. & ARTS 125, 130 (2010).

In contrast to focal points, encouragement has no effect on the choice of users to participate in the first place.[208] A focal point can function as encouragement,[209] but the opposite is not true. Intermediaries that encourage users to participate in a conversation take a direct position regarding the types of content they welcome on their platform.[210] Thus, this type of nudge differs from the indirect, nontransparent strategy of channeling and leading nudges, which push users through aspects of the platform's design without their awareness.

Similar to the two previously discussed categories of nudges, an intermediary uses encouragement to influence the dissemination of content on both micro and macro levels. The intermediary encourages individuals to generate and consume specific types of content.[211] It also motivates social dynamics and increases the likelihood for crossing the threshold to dissemination.[212] When an individual user sees in his newsfeed that his friends adopted a specific type of content by liking and sharing it, the likelihood for him to cross the threshold and act in the same way increases.[213] Utilizing the social network may lead to "mass interpersonal persuasion."[214] Consequently, the proportion of content the intermediary aims to promote increases.[215]

---

208. *Compare Dyroff*, 2017 WL 5665670, at *2 (encouragement), *and* Vision Sec., LLC v. Xcentric Ventures, LLC, No. 2:13-CV-00926, 2015 WL 12780892, at *2 (D. Utah Aug. 27, 2015) (encouragement), *and* McNeill, *supra* note 155 (encouragement), *with Terms of Service*, THEDIRTY.COM, https://thedirty.com/terms-of-service/ [https://perma.cc/9J4S-U7T6] (last visited Aug. 27, 2018) (focal point).

209. A focal point such as "Dirty World" also constitutes an encouragement to speech tort. However, a slogan that encourages tortious content such as "Keep it Juicy" is not a focal point because it does not influence the preliminary decision to participate and does not influence the social network's composition. *See* McNeill, *supra* note 155; *Terms of Service*, *supra* note 208.

210. *See Terms of Service*, *supra* note 208. The person who has been nudged by encouragements to create specific types of content is usually aware of being nudged. *See* McNeill, *supra* note 155.

211. *See* McNeill, *supra* note 155; *Terms of Service*, *supra* note 208.

212. *See* McNeill, *supra* note 155 ("The [w]eb site – a national message board that *urges* its anonymous collegiate visitors to post salacious gossip about their classmates." (emphasis added)).

213. *See* Zittrain, *supra* note 69, at 336. It should be noted that in many cases, intermediaries utilize algorithms to prioritize newsfeed content created by a user's close friends and family, which reinforces existing biases and further encourages dissemination. *See* SUNSTEIN, *supra* note 139, at 16.

214. *See* B.J. FOGG, MASS INTERPERSONAL PERSUASION: AN EARLY VIEW OF A NEW PHENOMENON 23, 24 (2008), http://captology.stanford.edu/wp-content/uploads/2014/03/MIP_Fogg_Stanford.pdf [https://perma.cc/89PP-VJ26].

215. *See id.* at 33; KADUSHIN, *supra* note 63, at 146.

a. Nuances of Encouragement and Gravity of Harm

Intermediaries may encourage defamatory[216] or negative content[217] in explicit and implicit ways. They can address all participants in general, and encourage them to generate specific types of content through slogans and banners.[218] They can also use more innovative strategies of encouragement to influence users in more profound ways.[219] To do so, they use studies and experiments on network structures and the flow of information.[220] These studies allow the identification of influential or susceptible hubs, which are central to dissemination.[221] Intermediaries personalize their encouragements and contact these hubs directly.[222] They can use smart chat-bots that are active in many platforms,[223] or send links to defamatory content or petitions to boycott a business to specific users and encourage them to sign on.[224] Similarly, intermediaries can encourage an "influential" user to comment on defamatory content and endorse it,[225] leading to

---

216. *See Fair Housing Council v. Roommates.com, LLC*, 489 F.3d 921, 928 (9th Cir. 2007) (describing a hypothetical website, "[H]arrassthem.com," which contains the slogan "Don't Get Mad, Get Even").

217. For an example of a website encouraging negative content, see RIPOFF REPORT, *supra* note 201 ("Don't let them get away with it! Let the truth be known!"). *See also* McNeill, *supra* note 155.

218. *See, e.g.*, Poole, *supra* note 204, at 181–82.

219. Banners and slogans have limited influence. In fact, many users consider them nuisances. *See* ADAM L. PENENBERG, VIRAL LOOP: FROM FACEBOOK TO TWITTER, HOW TODAY'S SMARTEST BUSINESSES GROW THEMSELVES 218, 222 (2009).

220. Intermediaries can receive information on network structures by conducting studies. *See* Aral & Walker, *supra* note 69, at 337; Grimmelmann, *supra* note 66, at 223; Zittrain, *supra* note 69, at 336.

221. *See* KADUSHIN, *supra* note 63, at 143–45; Aral & Walker, *supra* note 69, at 337.

222. *See* Dyroff v. Ultimate Software Grp., Inc., No. 17-cv-05359-LB, 2017 WL 5665670, at *2 (N.D. Cal. Nov. 26, 2017) (finding that data mining and machine learning allowed intermediaries to personalize recommendations to users on content and discussion groups on the website); Whigham, *supra* note 58 (describing how Facebook can detect and influence susceptible minors).

223. *See* VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK 29 (2014); HARTZOG, *supra* note 13, at 202 ("[P]recision advertising can be used to exploit biases and perpetuate falsehoods in significantly corrosive ways . . . ."); TUROW, *supra* note 60, at 159. These strategies were used to promote election campaigns in the United States. *Compare* sources cited *supra* note 69, *with* Philip N. Howard et al., *Algorithms, Bots, and Political Communication in the US 2016 Election: The Challenge of Automated Political Communication for Election Law and Administration*, 15 J. INFO. TECH. & POL. 81, 83 (2018) (describing ways political actors can influence elections with bots).

224. *See* GREENFIELD, *supra* note 23, at 116; B.J. Fogg & Clifford Nass, *Silicon Sycophants: The Effects of Computers that Flatter*, 46 INT'L J. HUMAN COMPUTER STUD. 551, 552 (1997). Intermediaries and other stakeholders can influence election processes in a similar manner. *See* Zittrain, *supra* note 69, at 336 (describing the election experiment that applied enhanced influence to vote on specific users); *see also* Levi, *supra* note 19 (manuscript at 25–26).

225. NAHON & HEMSLEY, *supra* note 77, at 142. Influencing central hubs in a social network to spread specific types of content shifts context. Consequently, the content disseminated is perceived as socially authentic and more credible. Laura E. Bladow, Note, *Worth the Click: Why*

mass "word of mouth" dissemination throughout the social network.[226] Advanced technologies thus allow intermediaries to create efficient encouragement nudges as they utilize user data to personalize messages. This personalized targeting can result in deeper influence on individuals, social dynamics, and flows of information throughout the network—all of which are more effective than general slogans.[227]

Encouragement increases the likelihood of reaching individual and collective thresholds.[228] Consequently, the severity of harm, which increases and the likelihood for private ordering may also be impaired. Informational and reputational cascades will reduce the likelihood for victims and other participants to counter speech and clear their names. The extent of encouragement depends largely on the degree of a nudge.

E. Interim Summary

This Part demonstrates how intermediaries can influence social contexts by using various strategies and technologies. Social relationships influence the flow of information from the bottom up; however, intermediaries also nudge social dynamics and influence decision-making from the top down.[229] Intermediaries use design to generate focal points and consequently influence the context of their platforms and the identity and composition of their users.[230] They can also influence the rate and strength of particular types of content in less transparent ways by channeling and leading users to their desired choices.[231] In addition, they engage in encouragement strategies and influence the flow of information in a clear and direct manner.[232]

Every choice of architecture is unavoidably context-based, and there is no such thing as a completely "neutral" design.[233] Thus, the taxonomy set forth by this Article illustrates that intermediaries' choice architecture is not arbitrary and can actively encourage tortious

---

Greater FTC Enforcement Is Needed to Curtail Deceptive Practices in Influence Marketing, 59 WM. & MARY L. REV. 1123, 1151 (2018).

226. On social spreading, see NAHON & HEMSLEY, *supra* note 77, at 142.

227. *See* Balkin, *supra* note 54, at 1184 (stating that advanced technologies of big data and AI increased intermediaries' influences on users and third parties); Zuboff, *supra* note 194, at 85; Balkin, *supra* note 19 ("[D]igital companies collect enormous amounts of data about their end-users, and use this data to predict and control what end-users will do . . . .").

228. *See* Dyroff v. Ultimate Software Grp., Inc., No. 17-cv-05359-LB, 2017 WL 5665670, at *2 (N.D. Cal. Nov. 26, 2017); *see also* Zuboff, *supra* note 195, at 85.

229. *See* Aral & Walker, *supra* note 69, at 337; Grimmelmann, *supra* note 66, at 223; Whigham, *supra* note 58.

230. *See supra* Section II.D.1.

231. *See supra* Section II.D.2.

232. *See supra* Section II.D.3.

233. *See* THALER & SUNSTEIN, *supra* note 1, at 10–11.

content. Digital intermediaries have the means to arrange decision-making context. Their ability to influence the gravity of harm should alert lawmakers and policy makers to rethink the scope of intermediary liability.

Nudges are not uniform. Therefore, it would be inappropriate to evaluate liability for evil nudges according to one set of standards. Mapping and understanding central strategies of nudges and their influence on social context takes the first step towards assisting courts in accommodating just and efficient policy when determining an intermediary's liability.

Table 1. Summary of Central Influences of Online Nudges

	Focal Point	Channeling and Leading	Encouragement
The Nudge's Influence	▪ Framing the website in a specific context generates exclusive or inclusive mechanisms, which pushes specific homogenous individuals to participate. This influences the composition of participants, their behavior, and the type of content they generate. ▪ This nudge influences context at the stage of deciding to use the website.	▪ Priming participants step-by-step to generate and consume specific types of content. Channeling and leading nudges influence the type of content participants generate, consume, and disseminate. ▪This nudge influences decision-making contexts only after users have decided to use the platform. ▪ In contrast to focal point, this nudge is less transparent and less salient.	▪ Direct influences on social context. The intermediary increases the motivations of users to generate specific types of content. The intermediary can identify influential hubs and contact them directly, thus exacerbating word-of-mouth dissemination. ▪ This nudge influences the content users generate, but has no effect on their choice to participate in the first place. ▪ In contrast to channeling and leading, this nudge applies to users in direct ways.
Examples	▪ Focal point for tortious content: TheDirty.com. ▪ Focal Point for gossip: JuicyCampus.com. ▪ Focal points for negative content: BadBusiness.com,	▪ Defaults. ▪ Multiple-choice-menus that lead to extremism (Roommates.com). ▪ Tagging options that frame specific types of	▪ Banners and slogans such as: "Pure Evil," "Keep it juicy," "Don't let them get away with it," "Don't Get Mad—Get Even." ▪ Decentralized encouragement.

	Focal Point	Channeling and Leading	Encouragement
Examples, continued	Ripoff Report, and PissedConsumer.	content (con artist, rip-off). ▪ Designing filters and search engines using limited parameters to channel users to generate specific types of content.	▪ Detecting "influential" participants in the social network by using network analysis, big data, AI, and contacting specific "influential" hubs personally, sending links to defamatory content and encouraging them to "like" and spread the content. This maximizes the advantages of the social graph and generates the sociological process of mass interpersonal persuasion.
Influence on consumption and generation of specific types of content	▪ Focal points promote an inclusion/exclusion mechanism. Thus, they influence the participants' composition and increase their motivation to generate specific types of content at the individual level.	▪ Channeling and leading nudges influence choice by generating priming and framing effects and thus enhance the distribution and consumption of specific types of content.	▪ Encouragement nudges increase the likelihood that users generate and consume specific types of content.
Influences on social dynamics	▪ Focal point nudges lead to homogenous composition of participants. This affects the likelihood for informational and reputational cascades. It also leads to extremism and mass adoption and dissemination of the content that the focal point supports.	▪ Channeling and leading nudges increase generation and consumption of specific types of content at the individual level. The proliferation of such content increases the likelihood to cross thresholds for disseminating content in social dynamics, while generating reputational and informational cascades.	▪ Encouragement nudges directly influence users and motivate sociological dynamics of adoption and dissemination. Utilizing central hubs within networks maximizes influence through the social graph, and increases the repetition of content.

	Focal Point	Channeling and Leading	Encouragement
Nudges and the gravity of harm	▪ High rate of tortious, negative, extreme, or antisocial content. ▪ Homogeneity and similarity of "prior dispositions" decrease the likelihood of private ordering by participants. ▪ Low probability of restoring reputation.	▪ Priming and framing effects by channeling and leading nudges influence the dissemination of tortious or negative content to various degrees, enhancing the gravity of harm and undermining the likelihood of correction.	▪ Encouragement influences the magnitude of tortious or negative content. It influences individuals and leads to social dynamics of adoption and dissemination. It decreases the likelihood of correction.

III. INTERMEDIARY LIABILITY AND SPEECH TORTS: THE LAW, NORMATIVE ANALYSIS, AND A CALL FOR CHANGE

A. Comparative Perspective

How does the law deal with intermediaries and contributory liability to defamatory content? This Part provides a comparative overview of policy models governing contributory liability of online intermediaries that nudge the harmful exchanges of tortious content. The analysis focuses on the United States and Europe to demonstrate that, due to the different extents of protection granted to freedom of expression, each legal system has adopted a different approach to this issue.[234]

1. United States

In the United States, lawsuits against online intermediaries are usually blocked by section 230(c)(1) of the Communications Decency Act (CDA). It provides that "[n]o provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider."[235] Under the subsection entitled "Protection for 'Good Samaritan' blocking and screening of offensive material[,]" Congress declared that online intermediaries should not be treated as publishers for material they did

234. *See* Oreste Pollicino & Marco Bassini, *Free Speech, Defamation and the Limits to Freedom of Expression in the EU: A Comparative Analysis*, *in* RESEARCH HANDBOOK ON EU INTERNET LAW 508, 513 (Andrej Savin & Jan Trzaskowski eds., 2014).

235. Communications Decency Act, 47 U.S.C. § 230(c)(1) (2012).

not develop.[236] Thus, a defendant that provides a forum for communicating materials is not likely to be held responsible as a content provider.[237] Courts have interpreted section 230 broadly—thus, section 230 has repeatedly shielded web enterprises from lawsuits.[238] Some courts, however, have criticized such vast immunity and have tried to narrow the statute's scope.[239] Courts have also tried to sidestep the CDA's immunity by employing various legal doctrines, such as promissory estoppel[240] and failure to warn.[241] Despite these attempts to narrow the CDA's immunity, the overall immunity for online intermediaries remains broad.[242]

This overall immunity regime applies to secondary liability; however, if the intermediary is "responsible" in whole or in part for the "creation or development" of content, courts may find the intermediary liable as an information content provider.[243] Section 230 does not define "creation" or "development"; therefore, the line between the service

---

236. *Id.* Congress enacted Section 230(c)(2) to encourage intermediaries to screen harmful content. It requires intermediaries who screen content to do so in good faith. *Id.* For an overview concerning the fact that no intermediary has lost its immunity because it did not make a good faith filtering decision, see Eric Goldman, *Online User Account Termination and 47 U.S.C. § 230(c)(2)*, 2 U.C. IRVINE L. REV. 659, 665 (2012).

237. *See* § 230(b)(1)–(2); Gallardo, *supra* note 161, at 735–38. Congress thus sought to promote self-regulation and free speech. While doing so, it allowed vibrant internet enterprises to prosper.

238. *See, e.g.*, Nemet Chevrolet, Ltd. v. Consumeraffairs.com, Inc., 591 F.3d 250, 254 n.4 (4th Cir. 2009); Zeran v. Am. Online, Inc., 129 F.3d 327, 330 (4th Cir. 1997) ("By its plain language, § 230 creates a federal immunity to *any* cause of action that would make service providers liable for information originating with a third-party user of the service." (emphasis added)); Giordano v. Romeo, 76 So. 3d 1100, 1101–02 (Fla. Dist. Ct. App. 2011); Caraccioli v. Facebook, Inc., 167 F. Supp. 3d 1056, 1065 (N.D. Cal. 2016) (holding that intermediaries are immune to liability as distributers and not only as publishers—thus, immunity applies even when intermediaries have knowledge of defamatory content and do not remove that content); Glob. Royalties, Ltd. v. Xcentric Ventures, LLC, 544 F. Supp. 2d 929, 933 (D. Ariz. 2008); Anupam Chander, *How Law Made Silicon Valley*, 63 EMORY L.J. 639, 653 (2014).

239. For example, Justice Easterbrook provided an alternative interpretation to section 230—treating it as a definition clause rather than means for immunity. *See* Doe v. GTE Corp., 347 F.3d 655, 659 (7th Cir. 2003).

240. *See* Barnes v. Yahoo!, Inc., 570 F.3d 1096, 1108 (9th Cir. 2009). The court refused to hold Yahoo! liable pursuant to section 230. However, it held Yahoo! liable for promissory estoppel—a theory of recovery based on a breach of contract. *See* David S. Ardia, *Free Speech Savior or Shield for Scoundrels: An Empirical Study of Intermediary Immunity Under Section 230 of the Communications Decency Act*, 43 LOY. L.A. L. REV. 373, 466–67 (2010).

241. For the Ninth Circuit's outline of the failure to warn exception to section 230 immunity, see *Doe v. Internet Brands, Inc.*, 824 F.3d 846, 851 (9th Cir. 2016); *Beckman v. Match.com, LLC*, 668 F. App'x 759, 760 (9th Cir. 2016). Yet, it should be noted that this exception is not widely adopted. *See* Dyroff v. Ultimate Software Grp., Inc., No. 17-cv-05359-LB, 2017 WL 5665670, at *5 (N.D. Cal. Nov. 26, 2017).

242. *See, e.g.*, *Zeran*, 129 F.3d at 330; *Giordano*, 76 So. 3d at 1101–02; *Caraccioli*, 167 F. Supp. 3d at 1065.

243. *See* Anupam Chander & Uyen P. Le, *Free Speech*, 100 IOWA L. REV. 501, 514 (2014); Zak Franklin, Comment, *Justice for Revenge Porn Victims: Legal Theories to Overcome Claims of Civil Immunity by Operators of Revenge Porn Websites*, 102 CAL. L. REV. 1303, 1316–17 (2014).

itself and the creation of information is blurred and the scope of liability is ambiguous.[244]

Arguably, many intermediaries do in fact "develop content" through the nudges described above. Different intermediaries influence the social network and the context of information flows in various ways. They generate focal points, channel and lead users to specific choices, and encourage the generation of particular types of content.[245] These activities form the basis of the claim that the intermediaries "develop content."[246] A body of case law has discussed intermediary liability for these types of influences.[247] At the outset, courts were reluctant to impose liability for influencing context and applied immunity in nearly all cases.[248] However, a seminal case, *Fair Housing Council v. Roommates.com*, created confusion with respect to intermediary liability.[249] After this decision, courts raised doubts about whether it is appropriate to continue to apply section 230 broadly.[250] This confusion resulted in conflicting decisions.[251]

a. The Roomates.com *Case*

Roommates.com is a website that enables users to find roommates.[252] The existing design of the Roommates.com required users to fill out a personal profile and answer several questions—including the user's sex, sexual orientation, and whether or not they have children.[253] It also required users to express their preferences with respect to roommates on each of these issues.[254] The answers were chosen from drop-down menus.[255] An internal search engine allowed

244. *See* Communications Decency Act, 47 U.S.C. § 230 (2012); Ken S. Myers, *Wikimmunity: Fitting the Communication Decency Act to Wikipedia*, 20 HARV. J.L. & TECH. 163, 187–201 (2006).

245. *See supra* Section II.D.

246. Generally, these plaintiffs base their legal suits on claims that intermediary influences are content development. *See, e.g.*, Fair Hous. Council v. Roommates.com, LLC, 489 F.3d 921, 925–29 (9th Cir. 2007). For a similar argument, see Olivier Sylvain, *Intermediary Design Duties*, 50 CONN. L. REV. 203, 218 (2018).

247. The Article reviews seminal case law in the following Sections.

248. *See* Ardia, *supra* note 240, at 461–63.

249. *See* Seth Stern, Note, *Fair Housing and Online Free Speech Collide in* Fair Housing Council of San Fernando Valley v. Roommates.Com, 58 DE PAUL L. REV. 559, 577 (2009). *See generally Fair Hous. Council*, 489 F.3d.

250. *See* Catherine Tremble, Note, *Wild Westworld: Section 230 of the CDA and Social Networks' Use of Machine-Learning Algorithms*, 86 FORDHAM L. REV. 825, 856 (2017).

251. *See* Catherine Gellis, *2012 State of the Law Regarding Internet Intermediary Liability for User-Generated Content*, 68 BUS. LAW. 289, 299–301 (2012).

252. *See Fair Hous. Council*, 489 F.3d at 924.

253. *Id.*

254. *Id.*

255. *Id.*

users to search roommates while filtering unfit matches according to these criteria.[256] The website also included an open section of user comments.[257] The intermediary sent users periodical emails, which included only potential roommate matches.[258]

The Fair Housing Council (FHC) sued Roommates.com—alleging that the questions included in the drop-down menus, the internal search engine, the filtering service, and even the open comment section violated the federal Fair Housing Act (FHA) and led to discrimination.[259] The case was originally dismissed due to section 230 immunity.[260] In order to overcome section 230's immunity, the FHC argued that by conditioning participation in the service on reporting restricted information, Roommates.com was an information content developer within the meaning of the statute—not a passive conduit.[261] On appeal, the Ninth Circuit reversed the district court's decision, declining to grant Roommates.com immunity.[262] The court held that the intermediary provided a limited set of prepopulated discriminatory answers and required users to choose one. Accordingly, the court found that Roommates.com was an information content provider because of the site's questionnaires and answer choices.[263] This conduct made Roommates.com a developer rather than a mere "passive transmitter" of information.[264] The court also declined to grant immunity for the site's internal search engine and email mechanism because these components did not use neutral tools, but rather channeled the distribution of discriminatory content.[265] The court upheld the immunity, however, for materials posted in the open comment section.[266]

---

256. *Id.*

257. *Id.*

258. *Id.*

259. *See* Fair Housing Act, Pub. L. No. 90-284 (codified as amended at 42 U.S.C. § 3604(c) (2012)); Fair Hous. Council v. Roommate.com, LLC, CV 03-09386PA(RZX), 2004 WL 3799488, at *2 (C.D. Cal. Sept. 30, 2004).

260. *See Fair Housing Council*, 2004 WL 3799488, at *6.

261. *Fair Hous. Council*, 489 F.3d at 926.

262. *See* Fair Hous. Council v. Roommates.com, LLC, 521 F.3d 1157, 1175 (9th Cir. 2008).

263. *Id.* at 1164–65.

264. *Id.* at 1166 ("By requiring subscribers to provide the information as a condition of accessing its service, and by providing a limited set of pre-populated answers, Roommate becomes much more than a passive transmitter of information provided by others; it becomes the developer, at least in part, of that information.").

265. *Id.* at 1167; *see also Fair Hous. Council*, 489 F.3d at 929 ("By categorizing, channeling and limiting the distribution of users' profiles, Roommate provides an additional layer of information that it is 'responsible' at least 'in part' for creating or developing.").

266. *Fair Hous. Council*, 521 F.3d at 1173–74.

In its decision, the court referred to the material contribution of the illegality test,[267] where a defendant's own acts must materially contribute to the illegality of the internet message for immunity to fail.[268] The court concluded that an intermediary that uses *neutral tools* to carry out what may be unlawful or illicit searches does not amount to "development" for purposes of determining section 230 immunity and, thus, is not liable for said illegal content.[269] In contrast, the drop-down menus in this case led to development of illegal discriminatory content and, for that reason, the majority found that Roommates.com was responsible for the discriminatory content.[270]

In reaching this conclusion, the majority essentially recognized that channeling and leading nudges, as well as encouragement nudges, could expose the intermediary to liability.[271] The dissenting opinion takes a narrower view of what it means to "develop" information online.[272] According to the dissent, providing a drop-down menu does not alone constitute "creating" or "developing" information in itself.[273] Rather, the dissent urged courts to examine whether the topics in drop-down menus are directly unlawful—for example, if the inquiry is a statutory violation or includes a defamatory statement.[274]

A few years later, the Ninth Circuit ruled in another case and adopted a narrower construction that excludes roommate selection from the reach of the FHA.[275] Thus, the rationale for denying Roommates.com immunity may not apply anymore because discriminatory statements can be lawful in this context.[276] It remains unclear whether the previous decision barred Roommates.com from

---

267. *Id.* at 1167–68.

268. *Id.* ("[W]e interpret the term 'development' as referring not merely to augmenting the content generally, but to materially contributing to its alleged unlawfulness. In other words, a website helps to develop unlawful content, and thus falls within the exception to section 230, if it contributes materially to the alleged illegality of the conduct.").

269. *Id.* at 1169–72. (distinguishing between the facts of *Roommates.com* and other cases where intermediaries designed drop-down menus and used neutral tools). *See also* Carafano v. Metrosplash.com, Inc., 339 F.3d 1119, 1124 (9th Cir. 2003); Lindsey A. Datte, Note, *Chaperoning Love Online: Dating Liability and the Wavering Application of CDA § 230*, 20 CARDOZO J.L. & GENDER 769, 781 (2014); Mark D. Quist, Comment, *"Plumbing the Depths" of the CDA: Weighing the Competing Fourth and Seventh Circuit Standards of ISP Immunity Under Section 230 of the Communications Decency Act*, 20 GEO. MASON L. REV. 275, 297 (2012).

270. *See Fair Hous. Council*, 521 F.3d at 1172.

271. *Id.* at 1165–67.

272. *Id.* at 1176–1182 (McKeown, J., concurring in part and dissenting in part) ("The majority's unprecedented expansion of liability for Internet service providers threatens to chill the robust development of the Internet that Congress envisioned.").

273. *Id.* at 1182.

274. *See id.* at 1189.

275. *See* Fair Hous. Council v. Roommate.com, LLC, 666 F.3d 1216, 1223 (9th Cir. 2012).

276. *See id.* at 1222 ("Because we find that the FHA doesn't apply to the sharing of living units, it follows that it's not unlawful to discriminate in selecting a roommate.").

enjoying section 230 immunity due to general contribution to the creation of discriminatory content or because of the nature of the questions and filtering criteria themselves.[277] This case left four questions unanswered: (1) What are "neutral tools"?, (2) What is "development" of content?, (3) What is "material contribution" to illegality"?, and (4) Would a wider range of choices in drop-down menus lead courts to a different conclusion regarding the intermediary's contribution to discriminatory content?[278]

Legal scholars have debated this case and its implications. Some researchers advocated for the Ninth Circuit outcome—suggesting that courts should apply it to intermediaries that design platforms aimed at enhancing harmful content.[279] Recent scholarship advocating for this result suggests that this case allows victims of online torts to overcome the barrier of section 230 immunity when intermediaries structure, sort, and sometimes sell user data.[280] Other scholars contend that the outcome was desirable, but note that it was inconsistent with section 230 and previous case law.[281] Be that as it may, most scholars criticized the ambiguity that the decision created, which may chill interactive innovation.[282]

---

277. *See* JACQUELINE LIPTON, RETHINKING CYBERLAW: A NEW VISION FOR INTERNET LAW 136 (2015); Sylvain, *supra* note 246, at 262 ("We might understand the *Roommates* opinion to suggest that a provider cannot be immune when it has *knowingly* designed its service or application in order to elicit illegal third-party content. . . . As with most website developers, the company was probably very attentive to the substantive preference options from which it allowed users to choose, as well as the way it presented the choices for selection (i.e., choice architecture). But the *Roommates* court did not frame its opinion in this way.").

278. *See* Harman, *supra* note 193, at 160; Christian Kaiser, *Paying for Nude Celebrities: Testing the Outer Limits of Roommates.com, Accusearch, and Section 230 Immunity*, 11 WASH. J.L. TECH. & ARTS 125, 133 (2015); Lynn C. Percival, *Public Policy Favoritism in the Online World: Contract Voidability Meets the Communications Decency Act*, 17 TEX. WESLEYAN L. REV. 165, 173 (2010).

279. *See, e.g.*, CITRON, supra note 76, at 177; Michael Burke, Note, *Cracks in the Armor?: The Future of the Communications Decency Act and Potential Challenges to the Protections of Section 230 to Gossip Web Sites*, 17 B.U. J. SCI. & TECH. L. 232, 256 (2011); Zieglowsky, *supra* note 153, at 1320.

280. *See* Sylvain, *supra* note 246, at 271–72; Tremble, *supra* note 250, at 868.

281. *See* Molly Sachson, *The Big Bad Internet: Reassessing Service Provider Immunity Under §230 to Protect the Private Individual from Unrestrained Internet Communication*, 25 J.C.R. & ECON. DEV. 353, 376 (2011); Bradley M. Smyer, Note, *Interactive Computer Service Liability for User-Generated Content after Roommates.com*, 43 U. MICH. J.L. REFORM 811, 835–38 (2010); Stern, *supra* note 249, at 586.

282. *See, e.g.*, Varty Defterderian, Note, *"Fair Housing Council v. Roommates.com": A New Path for Section 230 Immunity*, 24 BERKELEY TECH. L.J. 563, 592 (2009); Jeff Kosseff, *The Gradual Erosion of the Law that Shaped the Internet: Section 230's Evolution Over Two Decades*, 18 COLUM. SCI. & TECH. L. REV. 1, 37 (2016); Stern, *supra* note 249, at 586–87.

b. After Roommates.com

After the *Roommates.com* decision, courts expressed doubts regarding the scope of section 230 immunity.[283] This question has been the subject of many contradictory judicial decisions.[284] In general, courts have been inclined to find that a defendant is not an information content provider—thus choosing to err on the side of immunity. However, some courts have challenged traditional interpretations of section 230.[285] Thus, the standards for excluding intermediaries from immunity remain unclear.

As the following Sections show, a body of case law has developed in respect to the design of platforms and encouragement of harmful content. Many lawsuits have been filed against customer review services and rating websites, general rating platforms (e.g., ConsumerAffairs.com), and platforms for complaints (e.g., RipoffReport.com, and BadBusiness.com).[286]

In *Nemet Chevrolet, Ltd. v. Consumeraffairs.com, Inc.*,[287] internet users posted false negative reviews of Nemet's business (an automotive marketing organization for selling and serving automobiles) on ConsumerAffairs.com.[288] Nemet filed an action and alleged that ConsumerAffairs.com solicited posts from users, put them in particular categories, and edited them.[289] The court held that ConsumerAffairs.com's behavior did not exclude it from section 230 immunity because, in contrast to the *Roommates.com* case, the intermediary did not develop or encourage illegal content.[290]

283. *See* Kosseff, *supra* note 282, at 22 ("My analysis demonstrates that the erosion that began with the 2008 *Roommates.com* decision has accelerated, to a point where platforms have little certainty that they will be immune from claims arising from user content.").

284. *See* Dyroff v. Ultimate Software Grp., Inc., No. 17-CV-05359-LB, 2017 WL 5665670, at *5 (N.D. Cal. Nov. 26, 2017); Daniel v. Armslist, LLC, 913 N.W.2d 211, 224 (Wis. Ct. App. 2018).

285. *See* Amanda L. Cecil, Note, *Taking Back the Internet: Imposing Civil Liability on Interactive Computer Services in an Attempt to Provide an Adequate Remedy to Victims of Nonconsensual Pornography*, 71 WASH. & LEE L. REV. 2513, 2546 (2014).

286. *See, e.g.*, Jeff Kosseff, *Defending Section 230: The Value of Intermediary Immunity*, 15 J. TECH. L. & POL'Y 123, 124, 143 (2010) (describing how RipoffReport.com and ConsumerAffairs.com allow consumers to post reviews, most of which are negative and accuse the businesses of perpetrating frauds).

287. *See generally* Nemet Chevrolet, Ltd. v. Consumeraffairs.com, Inc., 591 F.3d 250 (4th Cir. 2009).

288. *Id.* at 252.

289. *Id.* at 256–58.

290. *Id.* at 257–58 (rejecting intermediary liability for channeling and leading, as well as encouragement nudges, by emphasizing that "a website operator who does not 'encourage illegal content' or 'design' its 'website to require users to input illegal content' is 'immune' under § 230 of the CDA."); *see also* Kimzey v. Yelp! Inc., 836 F.3d 1263, 1270–71 (9th Cir. 2016).

Courts have also discussed intermediary liability in creating platforms for negative reviews.[291] In most of these cases, courts applied section 230 immunity, despite criticizing the intermediaries' business model.[292] Thus, various courts concluded that RipoffReport.com, an intermediary that generated a focal point for negative and defamatory consumer reviews, was immune from liability because it neither generated the content nor adopted it.[293]

Another case, *Global Royalties, Ltd. v. Xcentric Ventures, LLC*, discussed the liability of RipoffReport.com for allowing users to label their reviews with defamatory titles and thus channel and lead users to publish defamatory reviews.[294] Likewise, the court applied immunity, concluding that allowing users to label their reviews with defamatory titles does not make the intermediary responsible—in whole or in part—for the "creation or development" of content, as long as users have the autonomy to choose among proposed options and add titles according to their discretion.[295]

In *Dyroff v. Ultimate Software Group, Inc.*, another court went one step further in upholding an intermediary's immunity. Here, the intermediary used data mining, machine learning, and algorithms that allowed it to analyze users' data. Moreover, it used this information to personally channel users to participate in particular groups and

291. *See, e.g.*, Glob. Royalties, Ltd. v. Xcentric Ventures, LLC, 544 F. Supp. 2d 929, 931 (D. Ariz. 2008); MCW, Inc. v. Badbusinessbureau.com, LLC, No. CIV.A.3:02-CV-2727-G, 2004 WL 833595, at *8 (N.D. Tex. Apr. 19, 2004) (discussing potential liability for the intermediaries RipoffReport.com and BadBusiness.com).

292. *See, e.g.*, Seldon v. Magedson, No. CV-13-00072-PHX-DGC, 2014 WL 1456316, at *6 (D. Ariz. Apr. 15, 2014); *Glob. Royalties*, 544 F. Supp. 2d at 933.

293. *See Seldon*, 2014 WL 1456316, at *6; Torati v. Hodak, No. 155979/12, 2014 WL 2620345, at *3 (N.Y. App. Div. June 11, 2014) (concluding that the name "Ripoff Report" and the website's slogan "Don't let them get away with it! Let the truth be known!" do not constitute liability).

294. *See Glob. Royalties*, 544 F. Supp. 2d at 930.

295. *Id.* at 932 ("Defendants provided a list of categories from which Sullivan selected the title 'Con Artists' for his post. As in our order dismissing the original complaint, we conclude that this participation is insufficient as a matter of law to make defendants information content providers with respect to the postings."); *see also* GW Equity LLC v. Xcentric Ventures LLC, No. 3:07-CV-976-O, 2009 WL 62173, at *17–18 (N.D. Tex. Jan. 9, 2009); Whitney Info. Network, Inc. v. Xcentric Ventures, LLC, No. 204-CV-47-FTM-34SPC, 2008 WL 450095, at *12 (M.D. Fla. Feb. 15, 2008).

consume particular types of content.[296] Other courts have even found intermediaries immune when they encouraged defamation.[297]

However, exceptional cases exist. In *Daniel v. Armslist*, the website Armslist.com allowed potential buyers and sellers of firearms and ammunition to contact one another, either by clicking on a link within the website or by using the contact information provided by the other party through the website.[298] This design facilitated illegal firearms purchases, one of which was a firearm used in a lethal shooting.[299] The plaintiff alleged that the design and operational features of Armslist.com affirmatively "encouraged" transactions through which prohibited purchasers acquired firearms.[300] The court interpreted *Roommates.com* broadly and did not apply immunity to website design features that facilitated illegal firearms purchases, even though only some of the transactions' sales ended up being illegal on the buyer's side.[301]

---

296. *See* Dyroff v. Ultimate Software Grp., Inc., No. 17-CV-05359-LB, 2017 WL 5665670, at *1, *8–10 (N.D. Cal. Nov. 26, 2017). *Dyroff* considers how data mining and machine learning allowed the intermediary to personalize recommendations to users on content and discussion groups on the website, sometimes channeling and leading users to unlawful content. *See id.* at *8. Because the intermediary steered one of the users to a discussion group dedicated to the sale of narcotics, the user was able to buy heroin and died because he consumed it. *See id.* at *3–5. The court dismissed the case, ruling that recommendations to website users is an ordinary, neutral function of social networking websites. *See id.* at *1. The intermediary used neutral tools that merely provide a framework that could be utilized for proper or improper purposes. *See id.* at *1. As such, it did not "create" or "develop" the information even in part. *See id.*; *cf.* Tremble, *supra* note 250, at 866 (implying that cases like *Dyroff* could be compared to the email service in *Roommates.com* because both platforms gleaned new information from user content and behavior to create a site architecture that affected both mood and behavior).

297. *See Glob. Royalties*, 544 F. Supp. 2d at 933 ("It is obvious that a website entitled Ripoff Report encourages the publication of defamatory content. However, there is no authority for the proposition that this makes the website operator responsible, in whole or in part, for the 'creation or development' of every post on the site. Essentially, that is plaintiffs' position."); MCW, Inc. v. Badbusinessbureau.com, LLC, No. CIV.A.3:02-CV-2727-G, 2004 WL 833595, at *10 (N.D. Tex. Apr. 19, 2004) ("MCW alleges that the defendants actively encourage, instruct, and participate in the consumer complaints posted on the websites. Specifically, MCW contends, the defendants, in an e-mail signed by Magedson, encouraged a consumer to take photos of (1) the owner, (2) the owner's car with license plate, (3) the owner handing out Ripoff Reports in front of Haldane's offices, and (4) the Bernard Haldane sign in the background with the Ripoff Reports in hand, all so that the defendants could include these photos on the websites.").

298. Daniel v. Armslist, LLC, 913 N.W.2d 211, 215 (Wis. Ct. App. 2018).

299. *Id.* at 217.

300. *Id.* at 215–16 (summarizing Armslist's alleged misconduct as (1) facilitating private sales by allowing users to limit searches to private sellers; (2) failing to flag "criminal" or "illegal" content; (3) warning against illegality but failing to offer specific legal guidance; (4) encouraging user anonymity; and (5) enabling buyers to evade a state waiting period that required federally-licensed firearms dealers to wait 48 hours after receiving a response from the background check system before transferring the firearm).

301. *Id.* at 222; *see also* Daniel v. Armslist, LLC, No. 2017AP344 (Wis. filed Aug. 15, 2018) (notification of court order) (informing petitioners that the Wisconsin Supreme Court will review Daniel v. Armslist, LLC); Eric Goldman, *Wisconsin Appeals Court Blows Open Big Holes in Section 230–*

Jones v. Dirty World was a major legal battle with many revolutions before the Sixth Circuit granted immunity.[302] TheDirty.com is a focal point for defamatory content. The name of the site in and of itself invites postings of "dirt."[303] In addition, the site included a "submit dirt" button that encouraged gossip. It also added brief, nasty remarks and tags to users' posts and published the selected submissions.[304] These posts offended many individuals—including Sara Jones, who brought an action against TheDirty.com—alleging state tort claims of defamation, false light, and intentional inflection of emotional distress.[305] The lower courts did not grant the intermediary immunity because it developed, invited, and encouraged defamatory content.[306] However, on the appeal, the Sixth Circuit applied a different interpretation of section 230, and held that the district court erred in finding that the website operators were the "creators" or "developers" of the content at issue.[307] In doing so, the Sixth Circuit rejected the encouragement test[308] and adopted the "material contribution" test for determining whether a website operator is "responsible, in whole or in

---

Daniel v. Armslist, TECH. & MARKETING L. BLOG (Apr. 25, 2018), https://blog.ericgoldman.org/archives/2018/04/wisconsin-appeals-court-blows-open-big-holes-in-section-230-daniel-v-armslist.htm [https://perma.cc/RH58-G95H] (noting that the *Armslist* opinion does not detail the exact circumstances when its statutory reading would support a Section 230 defense). Design features may thus allow plaintiffs to bypass section 230 and result in judicial denial of motions to dismiss, even if the design is neutral to illegality. *Cf.* Harrington v. Airbnb, Inc., 3:17-cv-00558-YY, 2018 WL 5619329, at *5 (D. Or. Oct. 30, 2018) (requiring users to display their picture in their profile may violate discrimination law); Eric Goldman, *Racial Discrimination Lawsuit Against Airbnb Has the Potential to Change Online Marketplaces–Harrington v. Airbnb*, TECH. & MARKETING L. BLOG (Nov. 2, 2018), https://blog.ericgoldman.org/archives/2018/11/racial-discrimination-lawsuit-against-airbnb-has-the-potential-to-change-online-marketplaces-harrington-v-airbnb.htm [https://perma.cc/LW2X-AJHE].

302. Jones v. Dirty World Entm't Recordings LLC, 755 F.3d 398, 402 (6th Cir. 2014); *see* Elizabeth M. Jaffe, *Imposing a Duty in an Online World: Holding the Web Host Liable for Cyberbullying*, 35 HASTINGS COMM. & ENT. L.J. 277, 287 (2013).

303. *Jones*, 755 F.3d at 402; *see also* THE DIRTY, *supra* note 9.

304. *Jones*, 755 F.3d at 403, 416 (describing how Richie, the website's manager, responded to posts and published his own comments on the discussion subjects, such as "[w]hy are all high school teachers freaks in the sack?" in response to a user's post about Jones).

305. *Id.* at 405.

306. *See* Jones v. Dirty World Entm't Recordings, LLC, 965 F. Supp. 2d 818, 822 (E.D. Ky. 2013); Jones v. Dirty World Entm't Recordings, LLC, 840 F. Supp. 2d 1008, 1012–13 (E.D. Ky. 2012); *see also* Eric Goldman, *Should TheDirty Website Be Liable For Encouraging Users to Gossip?*, FORBES (Nov. 25, 2013), https://www.forbes.com/sites/ericgoldman/2013/11/25/should-thedirty-website-be-liable-for-encouraging-users-to-gossip/#1a2ffcbdb966 [https://perma.cc/36PZ-8GNL] ("TheDirty lost Section 230 protection because it 'invited and encouraged' defamatory content, as evidenced by its name ('TheDirty'), Richie's screening of user submissions, and Richie's snarky comments appended to the user submissions.").

307. *Jones*, 755 F.3d at 415.

308. *Id.* at 414–15 ("More importantly, an encouragement test would inflate the meaning of 'development' to the point of eclipsing the immunity from publisher-liability that Congress established.").

part, for the creation or development of tortious information."[309] This interpretation led the court to grant immunity.[310]

In a similar case, the District Court for the Western District of Missouri granted immunity to the defendant, noting that section 230 governs how user-generated content is handled. As such, the court held that the CDA should not interfere with a website's name.[311] The court also held that merely encouraging defamatory posts is insufficient to overcome section 230 immunity.[312] In the court's view, intermediaries are not liable for focal points and encouragement.[313] Although scholars criticized these cases as granting too much defense protection for bad faith moderation, immunity is usually applied in these situations.[314]

An exception to the broad applicability of section 230 immunity are the corporate advocacy programs that purport to provide assistance to businesses with negative complaints by investigating and resolving the posted complaints for large fees.[315] Although some courts granted immunity to an intermediary that encouraged users to post negative reviews and directly profited from removing defamatory content, many other cases[316] that involved these programs were not rejected in preliminary stages.[317]

---

309. *Id.* at 413, 415 ("An adoption or ratification theory, however, is not only inconsistent with the material contribution standard of 'development' but also abuses the concept of responsibility. A website operator cannot be responsible for what makes another party's statement actionable by commenting on that statement post hoc. To be sure, a website operator's previous comments on prior postings could encourage subsequent invidious postings, but that loose understanding of responsibility collapses into the encouragement measure of 'development,' which we reject.").

310. *See id.* at 414–15, 417. The court applied a narrow interpretation of the material contribution test and concluded that a website owner who intentionally encourages illegal third-party postings to which he adds his own comments is not a "creator" or "developer" of that content. *Id.*

311. *See* S.C. v. Dirty World, LLC, No. 11-CV-00392-DW, 2012 WL 3335284, at *5 (W.D. Mo. Mar. 12, 2012).

312. *See id.* at *4.

313. *See id.*

314. *See, e.g.*, Danielle Keats Citron & Benjamin Wittes, *The Problem Isn't Just Backpage: Revisiting Section 230 Immunity*, 2 GEO. L. TECH. REV. 453, 468–70 (2018); Laura Cannon, Comment, *Indecent Communications: Revenge Porn and Congressional Intent of § 230(c)*, 90 TUL. L. REV. 471, 483–84 (2015); James Grimmelmann, *The Virtues of Moderation*, 17 YALE J.L. & TECH. 42, 105 (2015).

315. *See Why Corporate Advocacy*, RIPOFF REPORT, https://www.ripoffreport.com/corporate-advocacy-program/why-corporate-advocacy [https://perma.cc/NA9Q-DH5S] (last visited Sept. 9, 2018).

316. *See* Glob. Royalties, Ltd. v. Xcentric Ventures, LLC, 544 F. Supp. 2d 929, 933 (D. Ariz. 2008) ("It is obvious that a website entitled Ripoff Report encourages the publication of defamatory content. However, there is no authority for the proposition that this makes the website operator responsible, in whole or in part, for the "creation or development" of every post on the site. Essentially, that is plaintiffs' position".).

317. *See, e.g.*, Icon Health & Fitness v. Consumer Affairs.com, No. 1:16-cv-00168-DBP, 2017 WL 2728413, at *12 (D. Utah June 23, 2017); Vision Sec., LLC v. Xcentric Ventures, LLC,

In the United States, the scope of intermediary liability for nudging offensive content remains unclear. The *Roommates.com* decision recognized these influences as "content development" and revoked the immunity of intermediaries that had been protected by section 230 up to that point.[318] This case took the first step towards imposing liability for evil nudges. After *Roommates.com*, judicial decisions are inconsistent.[319] Yet, most courts still choose to err on the side of granting immunity.[320]

2. Europe

In Europe, intermediary liability is governed by the European Parliament's E-Commerce Directive.[321] The Directive does not impose a general duty of care on intermediaries to monitor content on their websites and also insulates intermediaries from liability—provided they remain passive facilitators of content and react upon actual knowledge of specific illegal content.[322] This knowledge-based safe haven protects intermediaries whose role is "merely technical, automatic and passive," but does not shield intermediaries that play an active role in hosting the content.[323] The Directive is somewhat dated,[324] and its classification may no longer be comprehensive. Many

---

No. 2:13-cv-00926-CW-BCW, 2015 WL 12780892, at *3 (D. Utah Aug. 27, 2015); Vo Grp., LLC v. Opinion, Corp., No. 8758/11, at 10 (N.Y. App. Div. May 22, 2012) (Court did not preclude liability for conditioning removal of tortious content on paying fees).

318. Fair Hous. Council v. Roommates.com, LLC, 489 F.3d 921, 928 (9th Cir. 2007).

319. *See* Cecil, *supra* note 285, at 2546.

320. *See id.*

321. *See* Council Directive 2000/31/EC, 2000 O.J. (L 178) (EC) 8; Joris van Hoboken, *The Legal Space for Innovative Ordering: On the Need to Update Selection Intermediary Liability in the EU*, 13 INT'L J. COMM. L. & POL'Y 1, 6 (2009); Broder Kleinschmidt, *An International Comparison of ISP's Liabilities for Unlawful Third Party Content*, 18 INT'L J.L. & INFO. TECH. 332, 345–48 (2010); Sophie Stalla-Bourdillon, *Sometimes One is Not Enough! Securing Freedom of Expression, Encouraging Private Regulation or Subsidizing Internet Intermediaries or All Three at the Same Time: The Dilemma of Internet Intermediaries' Liability*, 7 J. INT'L COM. L. TECH. 154, 155 (2012).

322. Article 14 provides that intermediaries engaged in "hosting" are not liable unless they have actual knowledge of illegal statements or refuse to remove them upon knowledge. *See* Ronen Perry & Tal Zarsky, *Liability for Online Anonymous Speech: Comparative and Economic Analyses*, 5 J. EUR. TORT L. 205, 220 (2014).

323. *See* Joined Cases C-236 & C-238/08, Google France, S.A.R.L. & Google Inc. v. Louis Vuitton Malletier SA et al., 2010 E.C.R. I-2417 ("[I]n order to establish whether the liability of a referencing service provider may be limited under Article 14 of Directive 2000/31, it is necessary to examine whether the role played by that service provider is neutral, in the sense that its conduct is merely technical, automatic and passive, pointing to a lack of knowledge or control of the data which it stores."); Corey Omer, *Intermediary Liability for Harmful Speech: Lessons from Abroad*, 28 HARV. J.L. & TECH. 289, 313 (2014); Stalla-Bourdillon, *supra* note 320, at 158.

324. *See* Perry & Zarsky, *supra* note 322, at 220.

content providers may not be "hosts" at all.[325] In such cases, the Directive does not apply at all.[326]

It remains unclear when courts consider an intermediary to be passive. Different courts have ruled that the Directive shields only "neutral" intermediaries.[327] Outside the scope of the E-Commerce Directive safe haven, the potential liability of intermediaries is extensive.[328] Thus, in the case of *Delfi,*[329] the Estonian Supreme Court interpreted the Directive narrowly and found the site liable for defamatory comments posted about a famous Estonian business executive—even though it followed the "notice-and-takedown" practice.[330] The court held that Delfi could not benefit from the Directive's safe haven because it allowed anonymous comments and did not apply sufficient measurements to prevent harm to third parties.[331] Therefore, in the court's view, the site should be held liable like any other publisher.[332] Delfi filed a complaint against the decision to the European Court of Human Rights (ECHR), claiming that its right to freedom of expression was violated.[333] The first section of the ECHR upheld the Estonian Court's ruling and did not find a proportional interference with freedom of expression according to Article 10 of the European Convention on Human Rights.[334] The Grand Chamber confirmed this decision.[335]

The court acknowledged that Delfi's anonymous comment section was notorious for its defamatory content, a fact which may have contributed to the court's final decision.[336] This judgment raises serious questions about intermediary liability. Moreover, it generates

---

325. *See id.* at 222; Peggy Valcke & Marieke Lenaerts, *Who's Author, Editor and Publisher in User-Generated Content? Applying Traditional Media Concepts to UGC Providers*, 24 INT'L REV. L. COMPUTERS & TECH. 119, 126 (2010).

326. *See* Perry & Zarsky, *supra* note 322, at 221.

327. *See* Tamiz v. Google Inc. [2013] EWCA (Civ) 68 [16], [2012] QB 449 (Eng.); Joined Cases C-236 & C-238/08, Google France, S.A.R.L. & Google Inc. v. Louis Vuitton Malletier SA et al., 2010 E.C.R. I-2417.

328. *See* Valcke & Lenaerts, *supra* note 325, at 126.

329. *See generally* Delfi AS v. Estonia [GC], No. 64569/09, Eur. Ct. H.R. (2015).

330. *See id.* at 70–71; Perry & Zarsky, *supra* note 322, at 221.

331. *See* Perry & Zarsky, *supra* note 322, at 221.

332. *See id.* ("The court acknowledged that Delfi's comment section was notorious for its defaming content, a fact which might have contributed to the court's final decision. Most importantly, Delfi was found liable even though it applied a 'notice and takedown' process and thus complied with the abovementioned requirements of the EU Directive.").

333. *See id.*

334. *See* Delfi AS, No. 64569/09 at 62. The court applied a narrow interpretation for intermediaries' technical functions. Martin Husovec, *ECtHR Rules on Liability of ISPs as a Restriction of Freedom of Speech*, 9 J. INTELL. PROP. L. & PRAC. 108, 109 (2014).

335. *See* Delfi AS, No. 64569/09 at 62.

336. *See* Perry & Zarsky, *supra* note 322, at 221–22.

confusion regarding the distinctions between online "publishers" and mere intermediaries.[337] Post *Delfi*, in *Index.hu Zrt v. Hungary*, the ECHR reached a different conclusion—holding that imposing liability on the website was a violation of Article 10.[338] Yet, the court did not retreat from its previous conclusions in *Delfi*.[339] Rather, it differentiated the nature of the published comments from the comments in *Delfi*.[340] The court held that Hungarian courts precluded a proper balancing between the right to freedom of expression and the right to reputation.[341] However, this ruling is confined to the individual circumstances of this particular case.[342] In *Pihl v. Sweden*, the ECHR continued the line of reasoning from *Index.hu Zrt*—while also considering the balance between human rights, the type of speech posted by the user, and the type of intermediary.[343]

Courts apply the Directive on a case-by-case basis,[344] which creates legal uncertainty regarding the scope of liability.[345] In *Delfi*, the court imposed liability because of the design of the platform and lack of sufficient measurements of precautions.[346] In the same manner, nudges are arguably aspects of choice architecture that extend beyond mere hosting. Therefore, it is likely that the Directive will not shield an intermediary who nudges specific types of speech, even if it did not have actual knowledge of the offending speech on its platform.[347]

---

337. *See id.* at 222 ("It should come as no surprise that the Delfi decision generated substantial confusion as to the distinction between online 'publishers' and mere 'intermediaries' and the extent of legal protection that adherence to a 'notice and takedown' process provides.").

338. *See* Magyar Tartalomszolgáltatók Egyesülete & Index.hu Zrt v. Hungary, No. 22947/13, Eur. Ct. H.R. (4th Sec.) 21 (2016).

339. *Id.*

340. *Id.* at 15, 17. The ECHR differentiated the nature of the comments that were published from the comments in Delfi and noted that the article in *Index.hu Zrt v. Hungary* was a matter of public interest and did not provoke offensive comments. *Id.*

341. *Id.* at 17.

342. *Id.* at 25 (Kuris, J., concurring).

343. *See* Pihl v. Sweden, No. 74742/14, Eur. Ct. H.R. (3d Sec.) 7 (2017) (considering the type of speech that did not amount to hate speech and the type of the intermediary—a nonprofit blogger that removed the comment upon notice and held the application inadmissible).

344. *See* Valcke & Lenaerts, *supra* note 325, at 125, 129 (stating that a lack of clear standards of liability leads to inconsistency).

345. *See id.* Post *Delfi*, it is unclear whether intermediaries can benefit from the safe haven and, if not, what is the standard of liability (e.g., negligence, publishers' strict liability).

346. *See* Delfi AS v. Estonia [GC], No. 64569/09, Eur. Ct. H.R. 81 (2015).

347. *See* Matthias Leistner, *Structural Aspects of Secondary (Provider) Liability in Europe*, 9 J. INTELL. PROP. L. & PRAC. 75, 77 (2014); Sophie Stalla-Bourdillon, *Making Intermediary Internet Service Providers Participate in the Regulatory Process Through Tort Law*, 23 INTELL. REV. L. COMP. & TECH. 153, 161 (2009).

B. Normative Considerations for Liability

Intermediary liability rests on the junction of a few areas of law. It balances constitutional rights and tort considerations. In addition, the technological context of intermediary liability involves considering the influence of liability on the path of innovation. Finding the right balance between these interests is a difficult judgment call—albeit a crucial one.

1. Constitutional Balance and the Base of Speech Torts

The civil rights at stake in defamation law involve human dignity, reputational interests, and freedom of speech.[348] The law must balance between the victim's reputation, the offender's right to free speech, and also the intermediary's rights. On the one hand, liability for defamation protects the basic elements of a person's status, dignity, and reputation as a member of society.[349] On the other hand, the law must also consider the right of free speech as a guard against government censorship.[350] The United States provides stronger protection for freedom of speech than other Western democracies,[351] both in political and commercial speech contexts.[352]

Several courts and scholars have contemplated why free speech should receive special protection.[353] The first rationale supporting the importance of free speech is that it promotes individual autonomy.[354] It enables the self-determination of an individual to express himself by

348. *See* Daniel C. Taylor, *Libel Tourism: Protecting Authors and Preserving Comity*, 99 GEO. L.J. 189, 196 (2010) ("Free societies must strike a balance between the rights of uninhibited speech and the interests of individuals in their reputations.").

349. *See* Peter G. Danchin, *Defaming Muhammad: Dignity, Harm, and Incitement to Religious Hatred*, 2 DUKE F.L. & SOC. CHANGE 5, 17 (2010).

350. *See* NEIL RICHARDS, INTELLECTUAL PRIVACY: RETHINKING CIVIL LIBERTIES IN THE DIGITAL AGE 10 (2015) ("Courts have interpreted the First Amendment broadly to prevent the government from censoring our speech, pushing us directly for its content, or creating legal rules that allow us to be sued for speaking the truth.").

351. *See* Pollicino & Bassini, *supra* note 234, at 514 (demonstrating that in the United States, the freedom of speech is protected more than in the EU). The different balance between free speech and reputation is even more prominent in the digital context. *See id.*

352. *See* Sorrell v. IMS Health Inc., 564 U.S. 552, 583 (2011) (striking down a Vermont law prohibiting the sale for marketing purposes of physicians' prescription records without their permission on the grounds that the law was not "content neutral"); Jane R. Bambauer & Derek E. Bambauer, *Information Libertarianism*, 105 CAL. L. REV. 335, 338 (2017); Tamara R. Piety, *"A Necessary Cost of Freedom"? The Incoherence of Sorrell v. IMS*, 64 ALA. L. REV. 1, 4 (2012) ("Sorrell may mean that henceforth, in practice, if not formally, commercial speech will be treated as fully protected.").

353. *See, e.g.*, RICHARDS, *supra* note 350, at 10 (reviewing influential theories which lay out justifications for the right to free speech).

354. *See* Joseph Raz, *Free Expression and Personal Identification*, 11 OXFORD J. LEGAL STUD. 303, 311–16 (1991).

familiarizing the public at large with his ways of life, allowing his preferences to gain public recognition and acceptability, and letting him know that he is not alone because his experiences are known to others.[355] A second rationale for protecting free speech is the search for truth.[356] Free speech assures that every expression can enter the marketplace of ideas.[357] A third rationale is based on the understanding that free speech is crucial for maintaining democracy.[358] Freedom of speech is required to assure the effectiveness of the democratic process by informing the governed of the acts of government and guaranteeing that policy is reached intelligently.[359] Contemporary theories on democracy focus on protecting and promoting a democratic, participatory culture.[360] Freedom of speech is required to assure an individual's ability to participate in the production and distribution of culture.[361] This theory stresses both individual liberty and collective self-governance.[362]

The digital age has pushed freedom of expression to the forefront of debate, raising old policy concerns regarding expression.[363] In the digital age, intermediaries can easily influence social contexts, lead to harmful dynamics, and affect mass attacks by many individuals against a single victim.[364] Consequently, they exacerbate reputational harm—especially when intermediaries function as crowd leaders—by encouraging and influencing mob destructiveness,[365] or they design specific focal points that facilitate low-value, harmful speech.[366] One

355. *See id.*

356. *See* JOHN MILTON, AREOPAGITICA: A SPEECH FOR THE LIBERTY OF UNLICENSED PRINTING TO THE PARLIAMENT OF ENGLAND ¶ 2 (1644).

357. *See* Abrams v. United States, 250 U.S. 616, 630 (1919) ("[T]he best test for truth is the power of the thought to get itself accepted in the competition of the market, and that truth is the only ground upon which their wishes safely can be carried out.").

358. *See* ALEXANDER MEIKLEJOHN, FREE SPEECH AND ITS RELATION TO SELF-GOVERNMENT xii (Oxford Univ. Press 1965) (1948).

359. *See id.* at 8.

360. *See* Michael D. Birnhack, *More or Better? Shaping the Public Domain*, *in* THE FUTURE OF THE PUBLIC DOMAIN: IDENTIFYING THE COMMONS IN INFORMATION LAW 59, 71–72 (2006) ("This is the view that self-government in a democracy is composed not only of the momentary act of voting, but also of what happens in between elections."); Cass R. Sunstein, *Beyond the Republican Revival*, 97 YALE L.J. 1539, 1548–49, 1570 (1988).

361. *See* Birnhack, *supra* note 360, at 71.

362. *See id.* at 72.

363. *See id.* at 86.

364. *See* Brian Leiter, *Cleaning Cyber-Cesspools: Google and Free Speech*, *in* THE OFFENSIVE INTERNET: SPEECH, PRIVACY, AND REPUTATION 155, 163 (Saul Levmore & Martha Nussbaum eds., 2010).

365. *See* Danielle Keats Citron, *Civil Rights in Our Information*, *in* THE OFFENSIVE INTERNET: SPEECH, PRIVACY, AND REPUTATION 31, 48 (2010); CASS SUNSTEIN & REID HASTIE, WISER: GETTING BEYOND GROUPTHINK TO MAKE GROUPS SMARTER 22–24 (2014).

366. *See* Geoffrey R. Stone, *Privacy, the First Amendment and the Internet*, *in* THE OFFENSIVE INTERNET: SPEECH, PRIVACY, AND REPUTATION, *supra* note 364, at 174, 175.

may argue that the law should impose liability on intermediaries for influencing context. Accordingly, liability can be the key to mitigating harm and protecting civil rights of victims.

The liability regime governing cyberspace affects free speech.[367] Imposing liability on intermediaries for nudges may lead to less nudging and in turn result in a chilling effect on specific types of platforms and on the speech of some speakers who would hesitate before expressing themselves. The chill may extend to a lesser degree than the potential chill of host liability because the intermediary decides how to design the platforms.[368] Nevertheless, the concern remains. Imposing liability on influencers may chill the incentives to design online platforms devoted to nonconsensual topics and discussions. Such discussions could involve the criticizing of subjects that are not mainstream and platforms devoted for marginalized groups.[369] Such liability may chill gossip, which promotes intimacy in social relations and provides a bridge between communities, among other benefits.[370] However, it may also discourage complaints and negative speech that are important for democracy.

Liability for nudging may harm minority groups who are underrepresented in mainstream discussions. Nudges—particularly focal point nudges—may enhance minority legitimacy and encourage minorities to spread ideas that would have been suppressed otherwise due to fear of social objection.[371] Some nudges provide a signal to minorities that their nonconsensual speech is acceptable, allowing them to trust their audiences with possibly controversial ideas and enhance their involvement in the democratic process.[372] Nudges can enrich the market of ideas and enhance their autonomy.[373] Imposing liability on intermediaries for nudges may hinder these benefits.

---

367. *See* Chander & Le, *supra* note 243, at 506.

368. *See* Lavi, *supra* note 71, at 930. Chilling will result in different choice architecture and not on direct censorship of user speech from the platform. *See id.* at 878. On host liability for harmful user expressions, see *id.* In contrast to host liability, an intermediary controls a platform's design and can reduce its exposure to liability by avoiding actions that create a basis for inducement. *See* Felix T. Wu, *Collateral Censorship and the Limits of Intermediary Immunity*, 87 NOTRE DAME L. REV. 293, 344–45 (2011).

369. *See* Lavi, *supra* note 71, at 883.

370. On the benefits of gossip, see Zimmerman, *supra* note 156, at 333–34.

371. *See* Danielle Keats Citron & Helen Norton, *Intermediaries and Hate Speech: Fostering Digital Citizenship for Our Information Age*, 91 B.U. L. REV. 1435, 1445 (2011).

372. *See* Matthew Gentzkow & Jesse M. Shapiro, *Ideological Segregation Online and Offline*, 126 Q.J. ECON. 1799, 1799 (2011). Intermediary influence on context promotes speech that might not be expressed otherwise. This is especially relevant with regard to focal points, which allow ideological segregation. *See* Citron & Norton, *supra* note 371, at 1445.

373. *See* Shlomo Cohen, *Nudging and Informed Consent*, 13 AM. J. BIOETHICS 3, 9 (2013).

One may argue that some chilling that will be caused by exposing intermediaries that nudge speech torts to liability is desirable.[374] Blanket immunity would allow intermediaries to freely influence social contexts, enhance flows of offensive speech, and inflict severe reputational harm on unsuspecting users.[375] Immunity may also undermine the right to free speech itself. First, nudges may flame social dynamics and bring users to a "hot" state of mind.[376] Consequently, users may not think carefully about their choices, which could lead to the automatic spreading of offensive speech that they would not have spread otherwise.[377] They might regret publishing these expressions later.[378] Thus, nudges can undermine the autonomy of users. Second, an exemption from liability for evil nudges may push users to publish more falsehoods without accountability—accordingly intensifying their flow.[379] Thus, more weight would be ascribed to these falsehoods in respect to other expressions.[380] This may hinder the free competition of expression in the market of ideas and may undermine the search for the truth.[381] Third, nudges may impair democracy. They may encourage users to spread falsehoods and fake news about state officials, hinder the ability to reach an informed decision,[382] and even manipulate political expression.[383] Therefore, a degree of chill may be desirable and could strike the right balance between the benefits of free expression and the costs of its potential harm.

Another balance that must be struck is between speakers of harmful speech and the freedom of speech of their victims, considering the right to speak on both sides. Exempting intermediaries from

374. *See* SUNSTEIN, *supra* note 86, at 71.

375. *See* Lavi, *supra* note 61, at 166.

376. *See* THALER & SUNSTEIN, *supra* note 1, at 41.

377. Individuals in a hot state are generated by the intuitive system of thinking (system 1), in contrast to the analytic system (system 2). *See* KAHNEMAN, *supra* note 48, at 20.

378. *See* ALESSANDRO ACQUISTI ET AL., "I REGRETTED THE MINUTE I PRESSED SHARE": A QUALITATIVE STUDY OF REGRETS ON FACEBOOK 1 (2011); Yang Wang et al., *From Facebook Regrets to Facebook Privacy Nudges*, 74 OHIO ST. L.J. 1307, 1308 (2013) (explaining that individuals in a "hot" state regretted spreading offensive content in social networks in retrospect). On nudges that infringe autonomy, see T. M. Wilkinson, *Nudging and Manipulation*, 61 POL. STUD. 341, 344 (2013).

379. *See* CASS SUNSTEIN, CONSPIRACY THEORIES AND OTHER DANGEROUS IDEAS 26 (2014).

380. *See id.* (arguing that the more times individuals are exposed to a rumor, the more they tend to believe it); Pennycook et al., *supra* note 124, (manuscript at 2).

381. *See* SAUL LEVMORE & MARTHA C. NUSSBAUM, THE OFFENSIVE INTERNET: PRIVACY, SPEECH, AND REPUTATION 102 (2010).

382. *See* SUNSTEIN, *supra* note 86, at 9–11; VAIDHYANATHAN, *supra* note 68, at 171.

383. *See* Zittrain, *supra* note 69, at 335 (describing how the graphic sign of the friends that voted functioned as a nudge that encouraged people to vote). It should be noted that new technologies, and big data in particular, allow to target nudges more efficiently. *See, e.g.*, Levi, *supra* note 19 (manuscript at 26).

liability allows them to nudge with impunity.[384] This can encourage mass attacks directed at specific individuals.[385] These attacks may deny victims their ability to engage with others as equals, which might suppress a free public debate.[386] Exempting intermediaries from liability would not only impair the autonomy of victims, but also the free market of ideas and public participation.[387] The balancing act of this tort must therefore include the victim's freedom of expression and the constitutional rights related to both parties.[388]

The final balance that must be struck is between the intermediary rights to free speech and the rights of users and third parties. One may argue that imposing liability on intermediaries for nudging undermines their freedom to design platforms as they see fit—thus undermining their freedom of expression.[389] However, it might also be argued that nudges are not speech.[390] This is especially true in cases of channeling and leading nudges, which aim to aid user navigation within the platform.[391] Yet, one might still argue that the role of channeling and leading extends beyond a tool for navigation and expression of ideas.[392] As for focal points, the intermediary influences

384. *See* Jeremy K. Kessler and David E. Pozen, *The Search for an Egalitarian First Amendment*, 118 COLUM. L REV. 1953, 1994 (2018) ("[A]rguments involving *speech on both sides* focus on the degree to which one party's expressive activity compromises the ability of other private parties to exercise their own First Amendment rights." (emphasis in original)); Lavi, *supra* note 61, at 182.

385. *See id.* at 184.

386. *See* CITRON, *supra* note 76, at 5; SILVERMAN, *supra* note 15, at 80; TUFEKCI, *supra* note 124, at 179 (explaining that saying that a person's political view is stupid is free speech; yet, when a mass mob attack a political view, this may create fear and block free speech); Danielle Keats Citron & Benjamin Wittes, *The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity*, 86 FORDHAM L. REV. 401, 420 (2017) ("Individuals have difficulty expressing themselves in the face of online assaults.").

387. *See* SUNSTEIN, *supra* note 86, at 10–11.

388. *See* Andrew M. Koppelman, *Revenge Pornography and First Amendment Exceptions*, 65 EMORY L.J. 661, 675 (2016).

389. *See* Tim Wu, *Machine Speech*, 161 U. PA. L. REV. 1495, 1498 (2013). For related context that applies to sophisticated AI nudges, see Tony M. Massaro et al., *SIRI-OUSLY 2.0: What Artificial Intelligence Reveals About the First Amendment*, 101 MINN. L. REV. 2481, 2483 (2017) (suggesting ways in which the rise of AI may inspire critical engagement with free speech theory and doctrine).

390. *See* Wu, *supra* note 389, at 1517.

391. *See id.* at 1525. On the functionality doctrine, see generally *id.*

392. *See id.* at 1525–26 (referring to software navigation and map programs as harder cases of differentiation between communication of ideas and functionality, yet tending to believe that they are functional tools). Other scholars adopt a broad approach to free speech and argue that platforms direct users to material created by other and report it. *See* Eugene Volokh & Donald M. Falk, *First Amendment Protection for Search Engine Results* 3 (UCLA School of Law, Research Paper No. 12–22, 2012). Another approach is that algorithms represent the message of their developers and is tied to human editorial judgement. *See* Stuart Minor Benjamin, *Algorithms and Speech*, 161 U. PENN L REV. 1445, 1479 (2013); DAVID M. SKOVER AND RONALD K. L. COLLINS, ROBOTICA: SPEECH RIGHTS AND ARTIFICIAL INTELLIGENCE 35–37, 42 (2018) (explaining that for constitutional purposes, what really matters is that the receiver experiences speech—including

the content of conversation, and the choice architecture is not merely functional, but rather expressional.[393] The same goes with encouragement nudges, which are definitely understood as speech.[394]

Assuming nudges are speech, intermediaries cannot have it both ways.[395] They cannot claim to be active speakers when seeking First Amendment protection, and only navigational tools when facing tort liability.[396] By enjoying the right of free speech, they undermine their immunity claims from civil liability.[397]

2. Theories of Traditional Tort Law

a. Corrective Justice

A central justification for imposing liability is corrective justice. Aristotelian philosophy defines corrective justice as a rectification of harm—specifically, harm that was wrongfully caused by one person to another—by means of a direct transfer of resources from the injurer to the victim.[398] Accordingly, every particular interaction embodies correlative rights and duties that are imposed on both parties. This deontological, nonconsequentialist concept focuses on bilateral interactions, which are not reliant on external values.[399]

Corrective justice theorists offer different reasons and requirements for imposing the duty of rectification—including concepts

---

robotic speech—as meaningful and potentially useful and valuable); Part III of , ROBOTICA explains that the First Amendment should protect communications in all forms relevant to human utility.

393. *See* Wu, *supra* note 389, at 1519–21 (differentiating between communication of functional information and communication of ideas).

394. *See id.* at 1511–12. Based on the analysis of Tim Wu, there is no doubt that encouragements are acts of speech. *See id.*

395. *See* Oren Bracha & Frank Pasquale, *Federal Search Commission? Access, Fairness, and Accountability in the Law of Search*, 93 CORNELL L. REV. 1149, 1193 (2008). However, the courts reached different conclusions regarding search engines—recognizing intermediaries right of free speech for page-rank and rejecting their liability for optimization. *See* Langdon v. Google, Inc., 474 F. Supp. 2d 622, 630 (D. Del. 2007); Search King, Inc. v. Google Tech., Inc., No. CIV-02-1457-M, 2003 WL 21464568, at *3 (W.D. Okla. May 27, 2003). These rulings have been criticized in literature. *See* FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 164 (2015); Frank Pasquale, *Reforming the Law of Reputation*, 47 LOY. U. CHI. L.J. 515, 525 (2015); Wu, *supra* note 389, at 1498 (describing potential harms of computer-generated speech that invite regulation).

396. *See* Pasquale, *supra* note 395, at 524.

397. *See* RICHARDS, *supra* note 350, at 86.

398. *See* ARISTOTLE, NICOMACHEAN ETHICS 77 (W. D. Ross trans., 1999).

399. *See* Ernest J. Weinrib, *Correlativity, Personality, and the Emerging Consensus on Corrective Justice*, 2 THEORETICAL INQUIRIES L. 107, 110 (2001).

of faults and rights,[400] responsibility,[401] and non-reciprocal risk.[402] Most theorists explain that causation is insufficient for imposing liability.[403] As a result, negligence or moral fault must exist to justify compensation for the harm caused.[404]

The theory of nonreciprocal risks can explain why harm alone is insufficient for justifying liability.[405] Liability exists when a respondent generates a disproportionately excessive risk of harm, relative to the victim's risk-creating activity.[406] The entitlement to recover a loss is handed to all injured parties to the extent the risks imposed on them were nonreciprocal.[407] The goal is to distinguish between risks that violate individual interests and background risks that must be borne by society.[408]

In light of the bilateral correlative nature of torts, the literature on corrective justice tends to focus on "first order" liability of those who most directly and wrongfully caused an injury, and not on "second order" liability of third parties who are not direct tortfeasors.[409] However, intermediaries arguably create the framework for risks by allowing the activity and assisting it.[410] Therefore, they can be liable for the consequences alongside the direct wrongdoer, because the corrective justice concept is also feasible when several wrongdoers caused the harm.[411]

Thus, it is arguable that intermediaries who influence context by nudging speech torts actually cause harm. Their actions are more

400. *See* JULES L. COLMAN, RISKS AND WRONGS 645 (1992); Benjamin C. Zipursky, *Civil Recourse, Not Corrective Justice*, 91 GEO. L.J. 695, 718 (2003).

401. *See* Stephen R. Perry, *The Moral Foundations of Tort Law*, 77 IOWA L. REV. 449, 449 (1992). Weinrib points out that tort doctrine constructs the tort relationship, because liability treats the parties as doers and sufferers of the same injustice. *See* Ariel Porat, *Questioning the Idea of Correlativity in Weinrib's Theory of Corrective Justice*, 2 THEORETICAL INQUIRIES L. 161, 169 (2001); Weinrib, *supra* note 399, at 110.

402. *See* George P. Fletcher, *Fairness and Utility in Tort Theory*, 85 HARV. L. REV. 537, 537 (1972).

403. *See id.* at 562. *But see* Richard Epstein, *A Theory of Strict Liability*, 2 J. LEGAL STUD. 151, 157 (1973) (arguing that harm is sufficient to justify compensation). However, this theory of strict liability, which focuses on factual causality, has come under criticism. *See, e.g.*, Izhak Englard, *The System Builders: A Critical Appraisal of Modern American Tort Theory*, 9 J. LEGAL STUD. 27, 62 (1980).

404. *See* Englard, *supra* note 403, at 65.

405. *See* Fletcher, *supra* note 402, at 542.

406. *See id.*

407. *See id.*

408. *See id.* at 543.

409. *See, e.g.*, Englard, *supra* note 403, at 27.

410. *See* Ardia, *supra* note 240, at 393.

411. *See* Richard W. Wright, *Allocating Liability Among Multiple Responsible Causes: A Principled Defense of Joint and Several Liability for Actual Harm and Risk Exposure*, 21 U.C. DAVIS L. REV. 1141, 1160 (1988). In that case, every wrongdoer is liable to the plaintiff's damages and can claim subrogation from other wrongdoers. *See id.*

than a background risk and the reputational harm of the victim is their fault. Yet, a counter argument might suggest that nudging specific types of content differs from generating it. The users can choose whether to participate and publish defamatory content or to avoid publishing it altogether. It is difficult to determine whether, in the absence of nudges, users would avoid publishing offensive speech. Moreover, the intermediary's fault should not be taken for granted.

Justifying liability under the corrective justice theory depends on the extent of the nudge's influence on users' content.[412] Strong, explicit nudges are not merely a routine background risk and are not an inherent part of operating platforms—thus, intermediaries that generate them create a nonreciprocal risk and should bear liability as if they committed the speech tort themselves. In contrast, weak nudges do not maintain a causal link between the intermediary and the harm and, therefore, may not indicate fault. Thus, it is neither fair nor just to impose liability on intermediaries that generate weak nudges. In these cases, the user alone should be liable.

b. Efficiency

Efficiency is one of the central tenets of tort law—focusing on the maximization of wealth and the efficient allocation of risks.[413] In general, it does not account for deontological considerations.[414] According to this perspective, legal rules aim to incentivize efficient conduct ex ante and promote welfare maximization ex post.[415] In this regard, courts should not consider the harm to victims in isolation. Instead, courts should consider the costs and benefits of the activity, as well as the value that third parties gain when the activity is undertaken. Benefits in this analysis may include social values such as freedom of expression and innovation.

Scholarly literature usually deals with the economic analysis of direct liability, but shies away from discussing third-party liability.[416] However, expanding liability to third parties is required in the following cases: (1) when the enforcement of liability on the direct tortfeasor fails;[417] (2) when the third party can monitor and control the direct

412. *See* Lavi, *supra* note 61, at 184.

413. *See* Richard A. Posner & William M. Landes, *The Positive Economic Theory of Tort Law*, 15 GA. L. REV. 851, 851 (1980).

414. *See* Richard A. Posner, *The Ethical and Political Basis of Efficiency Norm in Common Law Adjudication*, 8 HOFSTRA L. REV. 487, 492 (1980).

415. *See* John R. Hicks, *The Foundations of Welfare Economics*, 49 ECON. J. 696, 708 (1939).

416. *See, e.g.*, Hamdani, *supra* note 21, at 56.

417. *See* Perry & Zarsky, *supra* note 322, at 207 (discussing the example of when the direct tortfeasor cannot be detected and unmasked).

wrongdoers;[418] (3) when sufficient incentives do not exist for private ordering;[419] and (4) when a legal rule can be applied at a reasonable cost.[420] While third-party liability is well established, legal scholarship has little to say about the standard by which this liability should attach to the third-party tortfeasor.[421]

In the case of online speech torts, enforcement failures might occur[422] because the direct offender might be anonymous and, even if he is identified, he might not have deep enough pockets to adequately compensate victims.[423] In addition, the intermediary's influence on the decision of the direct offenders to publish speech torts can hinder social ordering on the platform. Under such a circumstance, to whom should liability be allocated? Who is the cheapest cost avoider? This Subsection examines whether efficiency considerations support imposing liability on intermediaries, while considering the alternative of letting victims bear the damage. The analysis refers to three types of traditional costs associated with assigning liability: (1) primary cost of deterrence;[424] (2) secondary cost of loss spreading;[425] and (3) administrative litigation costs.[426]

In order to achieve maximum efficiency, liability should be allocated to the cheapest cost avoider. One may argue that imposing liability on intermediaries who push users to publish offensive content is efficient. Intermediaries who influence social context facilitate speech torts.[427] Some of them strongly push users to commit speech torts and even construct illicit markets for offensive speech and profit

---

418. *See* Reinier H. Kraakman, *Gatekeepers: The Anatomy of a Third-Party Enforcement Strategy*, 2 J.L. ECON. & ORG. 53, 56 (1986); Lavi, *supra* note 71, at 882; Douglas G. Lichtman & Eric Posner, *Holding Internet Service Providers Accountable*, 14 SUP. CT. ECON. REV. 221, 224 (2006).

419. *See* Kraakman, *supra* note 418, at 56; Lavi, *supra* note 71, at 882; Lichtman & Posner, *supra* note 418, at 224.

420. *See* Kraakman, *supra* note 418, at 56; Lavi, *supra* note 71, at 882; Lichtman & Posner, *supra* note 418, at 224.

421. *See* Hamdani, *supra* note 21, at 57 ("[L]ittle is known about the appropriate scope of third-party liability. Specifically, legal scholarship has little to say about the standard of liability that should apply to third parties.").

422. *See* Matthew Schruers, *The History and Economics of ISP Liability for Third Party Content*, 88 VA. L. REV. 205, 233 (2002).

423. *See* Perry & Zarsky, *supra* note 322, at 238.

424. *See* GUIDO CALABRESI, THE COSTS OF ACCIDENTS: A LEGAL AND ECONOMIC ANALYSIS 68 (1970).

425. *See id.* at 39. Secondary costs are the costs associated with bearing primary costs. Significant losses borne by one person are more likely to result in secondary losses (arising from the initial damage) than allocating a series of small losses to many people, or large sum of losses to deep-pocketed entities. *See id.*

426. *See id.* at 24.

427. *See* Deirdre K. Mulligan & Kenneth A. Bamberger, *Saving Governance-By-Design*, 106 CAL. L. REV. 697, 701 (2018).

from them.[428] Consequently, intermediaries have few market incentives to prevent defamation.[429] The intermediaries who nudge and influence users are the cheapest cost avoiders.[430] They control the nudges they create by design,[431] even when they nudge automatically by using algorithms.[432] Indeed, some technologies can lead to results that the intermediary cannot foresee ex ante.[433] Yet, the intermediary can choose the technology it implements.[434] Imposing liability on intermediaries will disincentivize them from utilizing evil nudges ex ante and will promote efficient deterrence.

Imposing liability could also incentivize intermediaries to mitigate harm caused by tortious speech.[435] Waiving intermediary liability from intermediaries, in fact, incentivizes them to nudge

---

428. On illicit markets, see Kraakman, *supra* note 418, at 66.

429. *See* Jaffe, *supra* note 302, at 281.

430. *See id.* at 283–84. This is because the intermediary is in control of the design of its platform, which tends to be nonneutral. *See* Omer Tene & Jules Polonetsky, *Taming the Golem: Challenges of Ethical Algorithmic Decision-Making*, 19 N.C. J.L. & TECH. 125, 136–37 (2017). In contrast to failing to remove harmful content, influencing context is not an omission. Thus, the exposure to liability depends on the intermediary's discretion and actions. *See id.* at 136. Consequently, the intermediary is the cheapest cost avoider relatively to the victim. *See* Jaffe, *supra* note 302, at 283–84.

431. *See* Mulligan & Bamberger, *supra* note 427, at 701 ("[D]esigning technology to 'bake in' values offers a seductively elegant and effective means of control."). See also in a related context of promoting privacy-by-design, Ari Ezra Waldman, *Privacy's Law of Design*, U.C. IRVINE L. REV. (forthcoming 2019) (manuscript at 5) ("[D]esign's awesome yet invisible capacity to manipulate those who exist inside its ecosystem requires us to consider the values we want design to promote.").

432. *See* Tene & Polonetsky, *supra* note 430, at 136. Intermediaries can control the parameters at the base of the algorithms ex ante. *See id.* at 138. On "policy neutral" vs. "policy directed" algorithms, see *id.* at 137–42; Jack M. Balkin, *The Three Laws of Robotics in the Age of Big Data*, 78 OHIO ST. L.J. 1217, 1224 (2017) ("When we criticize algorithms, we are really criticizing the programming, or the data, or their interaction. But equally important, we are also criticizing the use to which they are being put by the humans who programmed the algorithms, collected the data, or employed the algorithms and the data to perform particular tasks."). On government by design, see SKOVER & COLLINS *supra* note 392 at 27 (referring Apple's Siri that has her limitations by design: "[S]he sidesteps medical, legal, or spiritual counsel; she eschews criminal advice; and she prefers the precise and factual to the ambiguous and evaluative."); Mulligan & Bamberger, *supra* note 427, at 697.

433. *See* THIERER ET AL., *supra* note 115, at 31. For example, the intermediary does not always foresee the exact results of the use of Artificial intelligence and machine learning. *See, e.g.*, THIERER ET AL., *supra* note 115, at 31 ("[E]ven if the public could review them, the nature of machine-learning techniques can obviate the usefulness of review because the program is teaching itself."); Madeline Lamo & Ryan Calo, *Regulating Bot Speech*, 66 UCLA L. REV. (forthcoming 2019) ("Bots can also display emergent behavior, meaning behavior neither the programmer nor the user of the bot anticipated in advance.").

434. *See* Tene & Polonetsky, *supra* note 430, at 136; Matthew U. Scherer, *Of Wild Beasts and Digital Analogs: The Legal Status of Autonomous Systems*, 20 NEV. L. J. (forthcoming 2018) (manuscript at 36) (explaining that designers can impose limitations on the systems' culpabilities).

435. On this point, in a related context of copyright infringement, see Douglas Lichtman & William Landes, *Indirect Liability for Copyright Infringement: An Economic Perspective*, 16 HARV. J.L. & TECH. 395, 398 (2003).

irresponsibly and externalize the damage caused.[436] In addition, intermediaries normally have deeper pockets than individual victims, and are better suited to reduce secondary costs by bearing the loss themselves or by spreading it to all their users.[437] An increase in litigation costs is expected, but imposing liability on intermediaries is better than the alternative of leaving the victim without a remedy. This alternative will not bring about efficient deterrence and may impose heavy secondary costs on victims.[438] Even though defamation law protects reputations without requiring proof of financial or physical suffering,[439] the conclusion is still valid. This is because defamation harm may have economic consequences and can lead to costly psychological harm.[440] Therefore, the victim is not the best potential bearer of intangible harm.[441]

An in-depth examination reveals that efficiency considerations fail to provide answers regarding the allocation of liability when considering overall market characteristics. In this context, there are important costs in the market to be considered. Imposing liability on intermediaries might not be desirable because the utility achieved by improving deterrence in the relevant market may be lower than its costs.[442] Requiring intermediaries to compensate defamation victims may distort access to digital markets and hinder positive externalities generated by intermediaries.[443] Liability for focal point and encouragement nudges will probably discourage legitimate speech such as complaints and criticism. Holding intermediaries responsible may also deter channeling and leading nudges—thus stifling innovation and development of efficient web-navigation tools.[444]

Allocating liability to intermediaries may also increase secondary costs of loss spreading. Erroneous assessment of risks caused by nudges may lead intermediaries to disproportionately increase their service prices.[445] Not all intermediaries are born equal,

---

436. *See* Jaffe, *supra* note 302, at 288–89.

437. *See* Perry & Zarsky, *supra* note 322, at 239.

438. *See id.* at 231–32.

439. *See* Daniel J. Solove & Danielle Keats Citron, *Risk and Anxiety: A Theory of Data-Breach Harms*, 96 TEX. L. REV. 737, 768 (2018).

440. On anxiety as a cognizable harm in a related context, see *id.*

441. *See* Perry & Zarsky, *supra* note 322, at 239 (explaining that intermediaries can better spread costs through pricing and insurance than ordinary users).

442. *See, e.g.*, Kreimer, *supra* note 56, at 58; Jaffe, *supra* note 302, at 283–84.

443. *See* Kreimer, *supra* note 56, at 58.

444. *See* Stern, *supra* note 249, at 589–90 ("If all websites strictly followed the Ninth Circuit's guidance, the Internet will eventually resemble a gigantic library with no cataloging system.").

445. *See* Lichtman & Landes, *supra* note 435, at 398 (discussing the spreading of costs to copyright holders if equipment manufacturers were liable). When an intermediary has full

and not all have deep pockets. For example, it might be inefficient to impose liability on desirable, noncommercial intermediaries, especially if the liability for users' content that is a result of the nudge is *not* knowledge based. Such liability may cause these desirable intermediaries to turn away from the market or to refrain from investing in online platforms altogether.[446] Consequently, only large commercial intermediaries would prevail—thus limiting the choice between platforms and diversity of online markets and resulting in less competition.[447]

Furthermore, allocating liability to intermediaries would cause an increase in legal action and administrative costs. Deciding the question of liability is complex and involves interpretation.[448] Courts have to interpret whether the intermediary "created" or "developed" the offensive content, and in such cases, inquire into the question of the intermediary's fault.[449] Litigating such questions is costly and time consuming.[450] A case may undergo a number of procedural hurdles and take years to resolve.[451] The costs of running complex litigation may lead intermediaries to limit their activities to sub-optimal levels in order to reduce their exposure to liability.[452] The different considerations outlined in this Section make it difficult to assess the most efficient allocation of liability. Cost-benefit analysis leads to

---

information on the level of risks, he can efficiently spread the loss in a way that reflects the costs of liability. *See id.* at 404. By contrast, uncertainty regarding the risk would result in inefficient loss spreading that would in turn result in disproportional burden on users. *See* KENNETH A. BAMBERGER & DEIRDRE K. MULLIGAN, PRIVACY ON THE GROUND: DRIVING CORPORATE BEHAVIOR IN THE UNITED STATES AND EUROPE 244 (2015) (explaining that ambiguity regarding the exposure to liability leads business to adopt higher standards relatively to the standards that would have been adopted under clear rules).

446. Lavi, *supra* note 61, at 186.

447. *See id.*

448. *See* Citron & Wittes, *supra* note 386, at 423.

449. *See* Ardia, *supra* note 240, at 460.

450. *See* Matt C. Sanchez, Note, *The Web Difference: A Legal and Normative Rationale Against Liability for Online Reproduction of Third-Party Defamatory Content*, 22 HARV. J.L. & TECH. 301, 318 (2008).

451. For a related example, see *Viacom International, Inc. v. YouTube, Inc.*, 676 F.3d 19, 28 (2d Cir. 2012); Fiona Finlay-Hunt, Note, *Who's Leading the Blind? Aimster, Grokster, and Viacom's Vision of Knowledge in the New Digital Millennium*, 2013 COLUM. BUS. L. REV. 906, 924–33 (2013); Jonathan Stempel, *Google, Viacom Settle Landmark YouTube Lawsuit*, REUTERS (Mar. 18, 2014), https://www.reuters.com/article/us-google-viacom-lawsuit/google-viacom-settle-landmark-youtube-lawsuit-idUSBREA2H11220140318 [https://perma.cc/7SJQ-KRA6]. In the context of speech torts, the immunity of section 230 of the CDA does not allow lawsuits to advance beyond preliminary stages. *See* Note, *Section 230 as First Amendment Rule*, 131 HARV. L. REV. 2027, 2027 (2018). However, when plaintiffs bypass immunity by raising direct and contributory claims, complex litigation is prolonged. See the major law battle in *Jones v. Dirty World Entertainment Recordings, LLC*, 766 F. Supp. 2d 828, 831–36 (E.D. Ky. 2011).

452. *See* Lital Helman, *Pull Too Hard and the Rope May Break: On the Secondary Liability of Technology Providers for Copyright Infringement*, 19 TEX. INTELL. PROP. L.J. 111, 155 (2010).

different conclusions regarding the optimal liability standard and depends on the degree of influence on the social context.[453] The stronger the evil nudge is, the more benefits are gained by imposing liability.[454]

c. Efficiency and Technological Innovation

In the digital age, one cannot discuss the allocation of liability without considering technological innovation. Technology influences the flow of information by allowing for the development of filtering mechanisms, search engines, drop-down menus, and innovative applications. These applications facilitate access to information. Consequently, they enrich social life, the market of ideas, and democratic culture.[455] The liability regime stifles innovation and impacts its course.[456] The expected liability outcome influences investments in certain types of technologies and the adoption of business models.[457]

One may argue that an exemption from liability for nudging will enable freedom and openness, thereby incentivizing entrepreneurs to invest in technological ventures and digital markets. Consequently, they will develop many innovative platforms and applications, such as drop-down menus and filters that promote efficiency. Stricter liability, however, might stifle innovation.[458] It might impede the significant technological progress witnessed in recent years,[459] including increases

453. *See id.*

454. *See* Schruers, *supra* note 422, at 237–38. Strong nudges are clear and intermediaries may easily avoid them. *See* Alex Kozinski & Josh Goldfoot, *A Declaration of the Dependence of Cyberspace*, 32 COLUM. J.L. & ARTS 365, 367–68 (2009). The gravity of their harm is major. *See id.* In contrast, when the nudge is less powerful, the costs of allocating liability to the intermediary may exceed its benefits. *See id.* at 370.

455. *See* RAINIE & WELLMAN, *supra* note 51, at 256–63; RHEINGOLD, *supra* note 175, at 77–109.

456. *See* Gideon Parchomovsky & Alex Stein, *Torts and Innovation*, 107 MICH. L. REV. 285, 314 (2008). Evidence suggests that innovation thrives under liberal liability regimes. *See* Kyle Graham, *Of Frightened Horses and Autonomous Vehicles: Tort Law and its Assimilation of Innovations*, 52 SANTA CLARA L. REV. 1241, 1270 (2012); Parchomovsky & Stein, *supra*, at 314; Guy Pessach, *Deconstructing Disintermediation: A Skeptical Copyright Perspective*, 31 CARDOZO ARTS & ENT. L.J. 833, 864 (2013); Tal Zarsky, *The Privacy-Innovation Conundrum*, 19 LEWIS & CLARK L. REV. 115, 125–26 (2015).

457. *See* Dotan Oliar, *The Copyright-Innovation Tradeoff: Property Rules, Liability Rules, and Intentional Infliction of Harm*, 64 STAN. L. REV. 951, 1001 (2012); Pessach, *supra* note 456, at 864–65 (noting that YouTube's success was due to the copyright liability regime (notice-and-takedown)). Such a regime alone does not prevent the variety of popular copyrighted content that the intermediary hosted on the site. *See id.* at 864.

458. *See* THIERER ET AL., *supra* note 115, at 4 n.5. A similar argument arose in a related context of copyright infringement by concurring Judge Breyer in *MGM Studios Inc. v. Grokster, Ltd.*, 545 U.S. 913, 965–66 (2005) [hereinafter *Grokster*].

459. *See* Geoffrey A. Manne & Joshua D. Wright, *If Search Neutrality Is the Answer, What's the Question*, 2012 COLUM. BUS. L. REV. 151, 186 (2012); Stern, *supra* note 249, at 586–87, 590

in productivity and personal satisfaction.[460] Due to the ambiguity regarding the scope of liability, innovation could become too risky or expensive.[461]

Yet, liability likely would have a limited effect on innovation as long as it remains neutral to technologies and does not depend on the adoption of specific technologies.[462] Certainly some innovators will shy away from legally murky areas. Nevertheless, promoting innovation cannot be the sole justification for exempting intermediaries from the law.[463] There exists an even more fundamental reason why exemption from liability would be unwise. An overall immunity for all types of architecture designs will yield a generation of technology that facilitates behavior that our society has decided to prohibit.[464] Futhermore, it may disincentivize intermediaries from developing safer and more efficient technologies.[465] Likewise, anyone who conducts business of any complexity must discuss liability risks with legal counsel.[466] In many cases, innovation continues despite formidable legal regulations and ambiguity regarding the scope of liability.[467] Thus, the concern of impeding innovation might be over-stated.[468] In sum, imposing liability on nudges that online intermediaries form should not be ruled out. However, since there are different types of

---

(noting that imposing liability on intermediaries for the design will eventually bring the internet to resemble a library with no cataloging system).

460. *See* ANUPAM CHANDER, THE ELECTRONIC SILK ROAD: HOW THE WEB BINDS THE WORLD TOGETHER IN COMMERCE 84 (2013); Michael A. Carrier, *Copyright and Innovation: The Untold Story*, 2012 WIS. L. REV. 891, 941–42 (2012); Chander, *supra* note 238, at 690; Kozinski & Goldfoot, *supra* note 454, at 367.

461. *See, e.g.*, Kozinski & Goldfoot, *supra* note 454, at 367. For example, the use of machine learning makes it difficult for the intermediary to foresee the scope and degree of their nudges in advance. *See* THIERER ET AL., *supra* note 115, at 31. On learning algorithms, see VIKTOR MAYER-SCHONBERGER & THOMAS RAMGE, REINVENTING CAPITALISM IN THE AGE OF BIG DATA 98 (2018).

462. *See* Kim, *supra* note 11, at 394.

463. *See* HARTZOG, *supra* note 13, at 121 ("Companies should generally have the freedom to design technologies as they please, so long as they stay within particular thresholds, satisfy certain basic requirements like security and accuracy, and remain accountable for deceptive, abusive and dangerous design decisions."); Walman, *supra* note 431 (manuscript at 62) ("I am unwilling to surrender to the intellectual hegemony of innovation"); Kozinski & Goldfoot, *supra* note 454, at 371.

464. *See* Kozinski & Goldfoot, *supra* note 454, at 371. Judge Kozinski wrote this article following his decision on the *Roommates.com* case. *See id. See generally* Fair Hous. Council v. Roommates.com, LLC, 489 F.3d 921 (9th Cir. 2007).

465. *See* Danielle Keats Citron & Mary Anne Franks, *Criminalizing Revenge Porn*, 49 WAKE FOREST L. REV. 345, 390 (2014); Alanna Petroff, *Google, Microsoft Move to Block Child Porn*, CNN (Nov. 18, 2013, 9:10 AM), http://money.cnn.com/2013/11/18/technology/google-microsoft-child-porn/ [https://perma.cc/6UPD-EP5E] (referring to Google and Microsoft's recent efforts with regard to child pornography).

466. *See* Kozinski & Goldfoot, *supra* note 454, at 371.

467. *See id.*

468. *See id.*

nudges, a one-size-fits-all approach to intermediary liability is inappropriate.

C. Rethinking Intermediary Liability for Nudges

Intermediaries are not mere conduits.[469] As Section II.D demonstrates, intermediaries nudge users and influence speech.[470] The internet revolution allows nudges to influence context and information in many ways.[471] Consequently, the likelihood of causing severe harm increases significantly. How should the law respond to this harm? Should online intermediaries be liable for nudges? What is the appropriate standard of liability? Current laws do not provide clear answers to these questions.[472] This extensive ambiguity results in uncertainty and confusion. Moreover, it may strike an inappropriate balance between constitutional rights, lead to unjust and inefficient outcomes, and deter innovation. Intermediary liability attracts a great deal of scholarly attention.[473] Different scholars suggest guidelines regarding the scope of liability.[474] However, some of these suggestions are either over or underinclusive, while others are too ambiguous.[475]

1. Intermediary Liability: Scholarly Suggestions and Limitations

a. Active/Passive Test and the Level of Interaction with Content

An empirical study of section 230 case law identifies areas of judicial inquiry and justifications for excluding intermediaries from the immunity.[476] Judicial inquiry focuses on the role the intermediary plays in the creation of the content and seeks to assess whether the defendant was "responsible, in whole or in part, for the creation or

469. *See* Sylvain, *supra* note 246, at 218 (describing the design of platforms that collect, analyze, and sort user data for their own commercial reasons and arguing that these functions belie any suggestion that online intermediaries are merely passive conduits of user information).

470. *See supra* Section II.D.

471. *See id.*

472. See the contradicting results in courts' decisions, *supra* Section III.A.

473. *See* JOEL R. REIDENBERG ET AL., CTR. ON LAW & INFO. POLICY, FORDHAM LAW SCH., SECTION 230 OF THE COMMUNICATIONS DECENCY ACT: A SURVEY OF THE LEGAL LITERATURE AND REFORM PROPOSALS 8 (2012), https://www.fordham.edu/download/downloads/id/1825/clip_section_230_of_the_communications_decency_act_report.pdf [https://perma.cc/JHJ9-2FKD].

474. *See, e.g.*, Citron & Wittes, *supra* note 386, at 423; Sylvain, *supra* note 246, at 277.

475. *See* Delfi AS v. Estonia [GC], No. 64569/09, Eur. Ct. H.R. 23 (2015) (comparing intermediary liability to traditional gatekeepers and tending to hold them responsible for disseminations); Sanchez, *supra* note 450, at 317 (suggesting an overall immunity regime for all types of dissemination).

476. *See* Ardia, *supra* note 240, at 457–59.

development" of the harmful speech.[477] Courts tend to focus on several factors, including the degree to which the intermediary exercises editorial control over content, encourages the submission of illegal content, or facilitates the creation or publication of it.[478] Intermediaries interact with user-generated content in a wide spectrum from passive hosting to providing content.[479] The more significant the intermediary's interaction is, the greater the likelihood that courts will deny motions to dismiss against them.[480] The distinction between active and passive hosting[481] is reflected in the Ninth Circuit's decision in *Roommates.com*,[482] as well as in recent scholarly work.[483] However, this distinction is inconsistent with section 230, which does not differentiate active and passive intermediaries. Instead, section 230 shields both active and passive intermediaries.[484] It is also problematic because it incentivizes intermediaries to remain as passive as possible.[485] As a result of incentivizing passivity, intermediaries may refrain from designing beneficial, innovative systems.[486] Avoiding actively influencing content might allow spammers and scammers to take over the platform, which may significantly disrupt its use.[487]

b. Technological Architecture: Drop-Down Menus and Navigation Tools

Following the *Roommates.com* case, some commentators suggested a regulatory policy for intermediaries who prepopulate their platforms with drop-down menus and navigation tools.[488] Specifically,

---

477. *Id.* at 460 (quoting 47 U.S.C. § 230(f)(3) (2012)).

478. *See id.* at 461.

479. *See id.* (listing types of interactions with content such as passive host, linking to content, editing content, republication, manipulation, drop-down menus, and providing content).

480. *See id.* at 505–06.

481. *See id.*

482. *See* Fair Hous. Council v. Roommates.com, LLC, 521 F.3d 1157, 1161–76 (9th Cir. 2008).

483. *See, e.g.*, Sylvain, *supra* note 246, at 214 (suggesting that courts should shield providers from liability for third-party user online conduct only to the extent that such providers operate as true passive conduits, or actually take good faith steps to remove or block illegal content).

484. *See* Communications Decency Act, 47 U.S.C. § 230(c) (2012); Varty Defterderian, *Fair Housing Council v. Roommates.com: A New Path for Section 230 Immunity*, 24 BERKELEY TECH. L.J. 563, 573 (2009). Section 230 protects the "Good Samaritan" and immunizes from liability intermediaries that actively regulate their platforms (for example by screening content). Defterderian, *supra*, at 567.

485. *See* Ardia, *supra* note 240, at 505–06 (showing that judicial decisions reflect this distinction, and the more interactive an intermediary is, the more likely they are to be held liable).

486. *See* Lavi, *supra* note 61, at 211.

487. *See* Grimmelmann, *supra* note 314, at 62.

488. *See, e.g.*, Harman, *supra* note 193, at 171–74.

these commentators recommended the application of a conditional "notice-and-take-down" regime[489] to intermediaries that utilize the aforementioned design of platforms.[490]

This direct focus on navigation tools is also problematic and can discourage development of beneficial technologies and navigation tools that make it easier to find information. In addition, liability directed at particular technological design disincentives only channeling and leading nudges and leaves other nudges that are unrelated to navigation tools exempt from liability.

c. "Bad Faith" Intermediation

Many scholars suggest that the courts should impose liability on intermediaries who act in bad faith.[491] A determination of bad faith could involve an actor's level of intent when committing a given action. Danielle Keats Citron and others suggest that the "worst actors," such as extortion intermediaries,[492] revenge porn websites,[493] and intermediaries devoted to abuse, should be excluded from section 230 immunity.[494] Similarly, Nancy Kim suggests that courts should impose

---

489. *See id.* at 170. According to this regime, the intermediary benefits from a safe haven if he removes user generated content when notified that this content is suspected of being defamatory. *Id.* Namely, "[o]nly in the event that the content provider does not respond can it be found liable." *See* Perry & Zarsky, *supra* note 322, at 241.

490. *See* Perry & Zarsky, *supra* note 322, at 241.

491. *See* Citron & Wittes, *supra* note 386, at 416 ("Extending immunity to Bad Samaritans undermines § 230's mission by eliminating incentives for better behavior by those in the best position to minimize harm."); Citron & Wittes, *supra* note 314, at 468–70; Grimmelmann, *supra* note 314, at 105–06 (noting that allowing immunity to websites like The Dirty gives "too much deference for bad-faith moderation").

492. *See, e.g.*, CITRON, *supra* note 76, at 6. These intermediaries encourage users to submit gossip, defamation, mug shots, or nude photos and charge fees for their removal. *See* Vo Grp., LLC v. Opinion, Corp., No. 8758/11, at 1–2, 5 (N.Y. App. Div. May 22, 2012); CITRON, *supra* note 76, at 6.

493. *See* CITRON, *supra* note 76, at 7; Molly K. Land, *A Human Rights Perspective on U.S. Constitutional Protection of the Internet*, *in* THE INTERNET AND CONSTITUTIONAL LAW: THE PROTECTION OF FUNDAMENTAL RIGHTS AND CONSTITUTIONAL ADJUDICATION IN EUROPE 48, 68 (Graziella Romeo & Oreste Pollicino eds., 2016); Cecil, *supra* note 285, at 2551; Citron & Franks, *supra* note 465, at 389; Franklin, *supra* note 243, at 1306; Layla Goldnick, Note, *Coddling the Internet: How the CDA Exacerbates the Proliferation of Revenge Porn and Prevents the Meaningful Remedy for Its Victims*, 21 CARDOZO J.L. & GENDER 583, 627 (2015). For example, Hunter Moors' platform "Is Anyone Up," whose slogan is "pure evil," encouraged users to submit nude photos of their ex-spouses, harm their reputation, and humiliate them. Mary Anne Franks, *"Revenge Porn" Reform: A View from the Front Lines*, 69 FLA. L. REV. 1251, 1278, 1296 (2017).

494. *See* Citron & Wittes, *supra* note 386, at 416, 423 ("The courts should certainly not extend the CDA's safe harbor to Bad Samaritans. Instead, § 230(c)(1) should be read to apply only to Good Samaritans envisioned by its drafters: providers or users engaged in good faith efforts to restrict illegal activity, as was true of Prodigy."). It should be noted that this Article narrows the immunity more than the previous suggestions of Citron ("the safe harbor would be limited to providers or users that have taken reasonable steps to prevent or address the illegality of which plaintiffs are complaining.") *See id.* at 420.

liability on intermediaries whose business models are specifically intended to encourage behavior that is likely to result in online harassment or other harmful speech.[495] These intermediaries should be held accountable for the ill effects resulting from their underlying business models.[496]

These suggestions are a good starting point, but require more development. Exposing only the worst actors to liability is too narrow a solution. It allows many intermediaries that are not the worst actors to promote harmful content without responsibility, even if they exacerbate harm. Imposing liability on business models that cause or exacerbate harm does not clarify where to draw the line.[497] On one end of the spectrum, it is clear that the worst actors—intermediaries who condition participation on posting illegal content—should not be shielded from liability because they are the worst actors and their business models clearly result in ill effects.[498] Conversely, it is also clear that mere hosts should not be held liable.[499] However, a gray area exists regarding intermediaries that implicitly encourage speech torts,[500] yet are not the worst actors.[501] With this regard, some scholars propose a broader approach of conditioning section 230's shield from liability in taking reasonable steps to prevent unlawful uses of the platforms.[502]

Yet, the appropriate level of intent for holding the worst actors responsible under section 230 remains unclear. When section 230 immunity does not apply, should actors be held to a knowledge-based, negligence, or strict liability standard? Furthermore, what actions should be considered when adopting the intent standard?

---

495. *See* Nancy S. Kim, *Web Site Proprietorship and Online Harassment*, 2009 UTAH L. REV. 993, 993 (2009).

496. See *id.* at 1007, 1045. This suggestion includes platforms for gossip.

497. *See id.* at 1045.

498. *See* Doty, *supra* note 207, at 125.

499. *See* Communications Decency Act, 47 U.S.C. § 230(c)(1) (2012); Nemet Chevrolet, Ltd. v. Consumeraffairs.com, Inc., 591 F.3d 250, 260 (4th Cir. 2009); Zeran v. Am. Online, Inc., 129 F.3d 327, 334 (4th Cir. 1997); Caraccioli v. Facebook, Inc., 167 F. Supp. 3d 1056, 1065 (N.D. Cal. 2016); Glob. Royalties, Ltd. v. Xcentric Ventures, LLC, 544 F. Supp. 2d 929, 931 (D. Ariz. 2008); Giordano v. Romeo, 76 So. 3d 1100, 1102 (Fla. Dist. Ct. App. 2011).

500. *See* Section II.B.2.b. One example is nudging users to post gossip and complaints. *See id.*

501. *See* McDonald, *supra* note 152, at 271. Insights from the broken window social theory illustrate how small changes in context have extensive influence on behavior. *See* George L. Kelling & James Q. Wilson, *Broken Windows: The Police and Neighborhood Safety*, ATLANTIC (Mar. 1982), http://www.theatlantic.com/doc/198203/broken-windows [https://perma.cc/7Q6F-YJCS].

502. *See, e.g.*, Citron & Wittes, *supra* note 386, at 419.

d. Incentives of Speakers and Claims Directed at the Intermediaries' Own Acts

Felix Wu discusses collateral censorship that occurs when a private intermediary suppresses the speech of others in order to avoid liability that otherwise might be imposed as a result of that speech.[503] Collateral censorship stems from a disconnection between the incentives of intermediaries and the original speaker.[504] Intermediaries have different incentives to carry particular content than original speakers have to create it in the first place.[505] Felix Wu argues that "[t]hose incentives diverge both because original speakers obtain benefits from the speech not realized by intermediaries and because intermediaries face liability risks not borne by original speakers."[506] Applying the same law to intermediaries and original speakers alike, despite the divergence of incentives, would incentivize intermediaries "to suppress more speech than would be withheld by original speakers."[507] Intermediary immunity reacts to the issue of collateral censorship.[508] Yet, immunity is not the appropriate response to situations in which collateral censorship is not the problem. An intermediary who obtains social benefits from speech does not need the incentives that immunity provides to facilitate speech and the rationale for immunity diminishes.[509]

Felix Wu concludes that intermediary immunity should not apply to inducement claims because these claims tend to involve the intermediary's own direct acts. Inducement claims do not place the intermediary in the role of a speaker, and the incentives that immunity provides to facilitate speech are not needed for preventing collateral censorship.[510] An inducement claim is premised on showing "'clear

---

503. *See* Wu, *supra* note 368, at 295–96.

504. *See id.* at 296 ("The unique harm of collateral censorship, as opposed to self-censorship, lies in the incentives that intermediaries have to suppress more speech than would be withheld by original speakers. This additional suppression occurs because intermediaries have different incentives to carry particular content than original speakers have to create it in the first place.").

505. *See id.* at 296–97.

506. *Id.*

507. *Id.* at 296.

508. *Id.*

509. *Id.* at 331.

510. *Id.* at 344 ("Inducement claims are a type of claim to which intermediary immunity ought not to apply, because such claims are directly targeted at the intermediary's own acts, and do not place intermediaries in the role of speakers."). Courts have applied section 230 to immunize intermediaries from claims that are unrelated to the classification of publishers or speakers (such as aiding and abetting defamation). *See* Thomas D. Huycke, Note, *Licensed Anarchy: Anything Goes on the Internet? Revisiting the Boundaries of Section 230 Protection*, 111 W. VA. L. REV. 581, 592 (2009).

expression or other affirmative steps taken to foster unlawful speech"; thus, an intermediary seeking to avoid liability needs only to avoid affirmative acts that form the basis for an inducement.[511] In such cases, the intermediary has no incentive to engage in collateral censorship and the likelihood for a chilling effect is relatively low.[512]

Imposing liability on extreme cases is relatively easy; however, the normative question of liability remains unclear. Not extending immunity to intermediaries in claims that focus on intermediaries' own acts and do not treat them as publishers or speakers may also have an indirect chilling effect.[513] Even if exposure to liability in such cases will not result in over removal of users' content, it may open platform design and technology directly to litigation, leading to inefficiency.[514]

e. Fiduciary Intermediaries

Professor Jack Balkin applies the concept of fiduciaries to tackle the problem of online manipulation that data collection exacerbates.[515] According to this perspective—since digital companies collect vast amounts of user data and utilize this information to predict and influence user activity—intermediaries may be "the most important example of the new information fiduciaries of the digital age."[516] Under this concept, intermediaries should neither breach user trust[517] nor take actions that users would reasonably consider unexpected or abusive for digital companies to do. Fiduciary duties extend beyond an intermediary's written policies and include duties of good faith, respect, and nonmanipulation.[518] Balkin proposes that legal regulation can

---

511. Wu, *supra* note 368, at 344.

512. *See id.* at 344–45.

513. *See supra* Section III.B.1.

514. See the broad liability applied in *Delfi AS v. Estonia* [GC], No. 64569/09, Eur. Ct. H.R. (2015), which may lead to switching off reader comment sections, as some European websites have already done. *See* Paul McNally, *Guardian Digital Chief: Killing off Comments 'a Monumental Mistake'*, NEWSREWIRED (Feb. 3, 2015, 10:32 AM), https://www.newsrewired.com/2015/02/03/guardian-digital-chief-killing-off-comments-a-monumental-mistake/ [https://perma.cc/D7D2-EWX5].

515. *See* Balkin, *supra* note 19 (manuscript at 78).

516. *See* Jack M. Balkin, *The First Amendment in the Second Gilded Age*, BUFFALO L. REV. (forthcoming 2019) (manuscript at 31); Balkin, *supra* note 54, at 1160; Balkin, *supra* note 19 (manuscript at 68–69); Jack M. Balkin, *Information Fiduciaries and the First Amendment*, 49 U.C. DAVIS L. REV. 1183, 1229 (2016); Jack M. Balkin & Jonathan Zittrain, *A Grand Bargain to Make Tech Companies Trustworthy*, ATLANTIC (Oct. 3, 2016), https://www.theatlantic.com/technology/archive/2016/10/information-fiduciary/502346/ [https://perma.cc/FC9B-JR7M].

517. *See* ARI EZRA WALDMAN, PRIVACY AS TRUST: INFORMATION PRIVACY FOR AN INFORMATION AGE 87 (2018).

518. Balkin, *supra* note 19 (manuscript at 78).

manage the potential for conflicts of interest, so that intermediaries "will be able to monetize personal data in some ways but not others."[519]

Indeed, the idea of fiduciary intermediaries provides a good starting point. This solution, however, focuses on user data and breach of trust towards the user, and not third parties.[520] In addition, it does not consider the problem of data breaches of information held by these information fiduciaries. However, it can indirectly mitigate the problem of manipulative inducement by limiting its most influential strategies.[521] Fiduciary duties are likely to limit intermediaries from utilizing users' data through abusive strategies of nudging.[522] Imposing fiduciary duties can deter intermediaries from using manipulative influential strategies that involve targeted, personalized evil nudges.[523] Reducing the use of this strategy is expected to mitigate the harm of evil nudges.[524] Yet, this proposal is only complimentary to other possible solutions and it will not solve the problem of evil nudges altogether.

Existing suggestions for regulating intermediary liability are limited. Therefore, a more comprehensive framework is required. The following part suggests that negative influence caused by evil nudges can and should be subject to third-party liability. Following this analysis, this Article offers tailored guidelines for determining liability of intermediaries.

IV. FROM NUDGES TO INTERMEDIARY LIABILITY: A NEW FRAMEWORK

Online speech does not take place in a void, but rather in various contexts. Each context facilitates distinctive kinds of expressions, interactions, and activities among users.[525] As Part II describes, intermediaries can and do influence social networks by utilizing social structures, cognitive biases, and technologies.[526] Part II further outlines an innovative taxonomy of intermediary nudges and demonstrates how they influence user content, dissemination of said content, and the credibility ascribed to it. Due to the far-reaching

519. *See id.* (manuscript at 70).

520. *See* Balkin, *supra* note 54, at 1162. One should note that in another article, Balkin proposed the concept of nuisance to tackle data collection that influences third party opportunities. Yet, at this stage, this solution is underdeveloped. Moreover, it is unclear what would be considered a nuisance to third parties in the context of liability for speech torts, and what should be the standard of liability for nuisance. *See id.* at 1164.

521. *See* Balkin, *supra* note 19 (manuscript at 77–78).

522. *See* Balkin, *supra* note 516, at 1232.

523. *See* Balkin, *supra* note 19 (manuscript at 79).

524. *See* Balkin, *supra* note 516, at 1206–07.

525. *See* Lavi, *supra* note 71, at 894.

526. *See supra* Section II.D (taxonomy of online evil nudges).

influence of intermediaries, this Part advocates for the recognition of liability to nudges as part of tort law.

Scholarly work explores nudges in similar contexts. Researchers discuss uses of technologies for generating nudges, promoting behavioral changes, and enhancing user privacy.[527] Some studies focus on the power of technology in persuasion and the influence of intermediaries as social actors.[528] Others refer to the cognitive problems of internet users and propose that policy makers respond to these problems.[529] Yet, this Article is likely the first attempt to apply a descriptive social technological model, based on social contexts, to normative legal policy.

This Part focuses on social contexts as a central factor for determining intermediary liability and the standard of liability. Intermediaries influence social contexts in various ways by using different types of nudges.[530] They play a substantive role in encouraging or discouraging speech and social dynamics.[531] However, they influence context in every architecture choice, as there is no absolute "unbiased" context.[532] It is neither applicable nor desirable to amend inherent biases by changing policy in every situation. Therefore, differential nuanced liability regimes must be promulgated. This proposition strives to avoid a disproportionate chilling effect, while deterring intermediaries from facilitating offensive content.

The following Sections integrate sociological and behavioral insights on nudges with legal policy[533] and outline proportional negative incentives to nudges in speech tort. The focus should be on the nature of the nudge that pushes users to commit speech torts or, in other words, the act of nudging. It emphasizes deontological considerations of corrective justice but also corresponds with a consequential approach by taking into account efficiency considerations.

---

527. *See* Alessandro Acquisti & Jens Grossklags, *What Can Behavioral Economics Teach Us About Privacy*, *in* DIGITAL PRIVACY: THEORY, TECHNOLOGIES, AND PRACTICES 363, 369 (2007); LESLIE K. JOHN ET AL., THE BEST OF STRANGERS: CONTEXT-DEPENDENT WILLINGNESS TO DIVULGE PERSONAL INFORMATION 5 (2009); Alessandro Acquisti et al., *Privacy and Human Behavior in the Age of Information*, 347 SCI. 509, 511 (2015).

528. *See, e.g.*, FOGG, *supra* note 15, at 90; Howard et al., *supra* note 223, at 84 ("Most social bots are designed to operate over social media platforms, while pretending to be real human users."); Jones, *supra* note 115, at 164 (explaining the role of smart chatbots as social actors).

529. *See, e.g.*, Daniel J. Solove, *Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1880, 1883 (2013).

530. *See supra* Section II.D.

531. *See supra* Sections II.B–II.D.

532. *See* GILLESPIE, *supra* note 55, at 42; SUNSTEIN, *supra* note 1, at 118; THALER & SUNSTEIN, *supra* note 1, at 10–11; *supra* Section II.D.2.a.

533. See *supra* Part II, which reviewed these insights.

A. The Degree of Harmful Nudges

Different nudges cause different degrees of harm.[534] There are many different types of nudges ranging from marginal to strong.[535] The degree of influence a given nudge has on users should play a central role in outlining guidelines for determining intermediary liability. The type of nudge indicates its influence on social networks and explains the casual link between the influence and the speech tort committed by users.[536] The proposed framework applies differential standards of liability for different types of influence and does not rely on the liability-immunity dichotomy.

1. Differential Standards of Liability

In copyright infringement law, intermediary liability draws on a nuanced toolbox of liability. Moreover, copyright infringement law provides a natural starting point for this analysis. The following Subsections review these tools and propose adjustments to tailor them to nudges.

a. Lessons from Third-Party Liability in Copyright Infringement

Many intermediaries do not function as "mere hosts"[537]—thus, their liability should be contributory. The *Roommates.com* case recognized this fact.[538] It borrowed ideas from other fields of law, specifically contributory liability and inducement for copyright infringement,[539] though it did not explicitly mention the seminal cases in this context.[540] Yet, the analysis in *Roommates.com* neglects the

534. *See supra* Section II.D (referring to degrees of nudges in applying the framework).

535. *See supra* Section II.D.

536. *See infra* Section III.B.1.b.

537. On hosts liability, see Lavi, *supra* note 71, at 870–71.

538. *See* Defterderian, *supra* note 484, at 577.

539. *See id.* The emulating of copyright theories was reflected in the "material contribution to illegality" test. *See* Fair Hous. Council v. Roommates.com, LLC, 521 F.3d 1157, 1168–69 (9th Cir. 2008) ("A dating website that requires users to enter their sex, race, religion and marital status through drop-down menus, and that provides means for users to search along the same lines, retains its CDA immunity insofar as it does not contribute to any alleged illegality; this immunity is retained even if the website is sued for libel based on these characteristics because the website would not have contributed materially to any alleged defamation."); Defterderian, *supra* note 484, at 576.

540. *See* MGM Studios Inc. v. Grokster, Ltd., 545 U.S. 913, 914 (2005) (narrowing Sony's rule by allowing contributory liability to run in the presence of "clear expression or other affirmative steps taken to foster infringement," regardless of the substantial non-infringing uses); Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 442 (1984) (ruling that the manufacturers of home video recording devices, such as Betamax, cannot be liable for infringement

element of fault dictated by copyright law and may curb the activities of legitimate intermediaries.[541] A comprehensive discussion on theories of liability for copyright infringement is therefore required to outline a better model. Indeed, the objectives of defamation and copyright laws are different—thus, the laws are not completely analogous.[542] Adjusting the toolbox of liability standard in copyright to nudges, however, would refine the analysis and bring about a better policy.

b. Toolbox of Differential Standards

Three main doctrines govern third-party liability in copyright infringement. The first is vicarious liability, where a supervisory entity is held responsible for the activities of violators under its control.[543] This doctrine draws on agency principles and expands on them.[544] The second doctrine is contributory liability, in which a third party is liable when direct infringement takes place, it has knowledge of the activity, and contributes materially to the infringing conduct of the direct infringer.[545] This liability is based on participation in the activity or supplying the means for it.[546] In contrast to vicarious liability, contributory liability requires the defendant to have actual knowledge of the specific infringement.[547] The third doctrine is inducement—or the enticement to engage in an infringing activity.[548] Under this doctrine, the inducer "need not necessarily have the right and ability to

---

and holding that the test for contributory liability was whether a product "is capable of commercially significant non-infringing uses"); Defterderian, *supra* note 484, at 579–80.

541. *See* Defterderian, *supra* note 484, at 579–82. For example, the *Roommates.com* case may bring collateral censorship on platforms devoted to complaints. *See id.* at 582.

542. *See* Robert C. Post, *The Social Foundations of Defamation Law: Reputation and the Constitution*, 74 CAL. L. REV. 691, 727, 736 (1986). Defamation law is premised on autonomy and dignitary arguments whereas the traditional justification for intellectual property (IP) rights has been utilitarian. *See id.*; Mark A. Lemley, *Faith-Based Intellectual Property*, 62 UCLA L. REV. 1328, 1328, 1345 (2015).

543. *See* Lichtman & Landes, *supra* note 435, at 398.

544. *See* Fonovisa, Inc. v. Cherry Auction, 76 F.3d 259, 261–64 (9th Cir. 1996); Shapiro, Bernstein & Co. v H. L. Green Co., 316 F.2d 304, 307 (2d Cir. 1963); Dreamland Ball Room, Inc. v. Shapiro, Bernstein & Co., 36 F.2d 354, 355 (7th Cir. 1929); Dan Burk, *Toward an Epistemology of ISP Secondary Liability*, 24 PHIL. & TECH. 437, 439 (2011); Helman, *supra* note 452, at 116.

545. *See* Gershwin Publ'g Corp. v. Columbia Artists Mgmt., Inc., 443 F.2d 1159, 1162 (2d Cir. 1971).

546. *See* Burk, *supra* note 544, at 440 ("Here the indirect infringement stems not from supervision or control, but from either participation in the infringing enterprise, or from supplying the means to infringe, without actually committing any of the acts prohibited by the exclusive rights of the copyright owner.").

547. *See* Kalem Co. v. Harper Bros., 222 U.S. 55, 63 (1911); Helman, *supra* note 452, at 115; Daniel Kohler, *A Question of Intent: Why Inducement Liability Should Preclude Protection Under the Safe Harbor Provisions of the Digital Millennium Copyright Act*, 41 SW. L. REV. 487, 493 (2012).

548. *See* Burk, *supra* note 544, at 440.

control the violator."[549] Inducement necessarily requires a degree of knowledge of infringements in general, and an intent to encourage the infringing activity.[550] Some courts refer to this doctrine as a derivative of secondary liability,[551] while others refer to it as an independent doctrine.[552]

Third-party liability for copyright infringement online raises complex questions. Often, intermediaries provide tools that can facilitate infringement (e.g., file sharing software) and subsequently look the other way.[553] The doctrine of secondary liability provides that such behavior is unacceptable.[554] The Digital Copyright Millennium Act (DMCA) was enacted to strike a balance between online intermediaries and the interests of copyright holders.[555] This law did not outline new standards. Rather, the DCMA added a second stage for evaluating liability.[556] It includes a safe harbor, which provides intermediaries with a shield.[557] Intermediaries are exempt from some liability at a cost of fulfilling copyright enforcement duties[558] that require removal of infringing materials upon that intermediary obtaining knowledge of such material.[559]

Robust litigation revolves around the degree of knowledge needed for liability.[560] *Viacom International v. YouTube* created

---

549. *See id.*

550. *See id.* (explaining that the standard of inducement does not require actual knowledge of specific infringements—instead, it is enough to have general knowledge of infringements.); Kohler, *supra* note 547, at 495.

551. *See* Arista Records LLC v. Lime Grp. LLC, 784 F. Supp. 2d 398, 436 (S.D.N.Y. 2011) ("[I]nfringement claims based on secondary liability, including claims for inducement of infringement, derive from the common law." (citing MGM Studios Inc. v. Grokster, Ltd., 545 U.S. 913, 930, 934–36 (2005))).

552. *See* Columbia Pictures Indus., Inc. v. Fung, No. CV 06-5578 SVW(JCx), 2009 WL 6355911, at *7 (C.D. Cal. Dec. 21, 2009) ("The first two theories (material contribution and inducement) are known collectively as 'contributory liability.'"); Kohler, *supra* note 547, at 502 n.118.

553. *See, e.g.*, Kozinski & Goldfoot, *supra* note 454, at 367, 369.

554. *See id.* at 367.

555. *See* Digital Millennium Copyright Act, 17 U.S.C. § 512 (2018); ALS Scan, Inc. v. RemarQ Cmties., Inc., 239 F.3d 619, 625 (4th Cir. 2001).

556. *See* Jonathan J. Darrow & Gerald R. Ferrera, *Social Networking Web Sites and the DMCA: A Safe-Harbor from Copyright Infringement Liability or the Perfect Storm?*, 6 NW. J. TECH. & INTELL. PROP. 1, 26 (2007).

557. *See id.* at 12.

558. *See* Niva Elkin-Koren, *Making Technology Visible: Liability of Internet Service Providers for Peer-to-Peer Traffic*, 9 N.Y.U. J. LEGIS. & PUB. POL'Y 15, 17 (2005).

559. *See* Digital Millennium Copyright Act, 17 U.S.C. § 512(c)(1) (2012); Amir Hassanabadi, Note, *Viacom v. YouTube – All Eyes Blind: The Limits of the DMCA in a Web 2.0 World*, 26 BERKELEY TECH. L.J. 405, 412–13 (2011).

560. *See* Viacom Int'l Inc. v. YouTube, Inc., 718 F. Supp. 2d 514, 518 (S.D.N.Y. 2010). There are different levels of knowledge: actual knowledge, objective, "red flag," "willful blindness," or intent. *See id.* at 516–17, 520; Tamlin H. Bason, *Court Affirms DMCA Red Flag Standard, Recognizes Willful Blindness Liability Doctrine*, BLOOMBERG: BNA (Apr. 6, 2012),

ambiguity in this regard.[561] The US Court of Appeals for the Second Circuit did not require the indirect infringer to have actual knowledge of the infringement and was satisfied with objective knowledge that was articulated in the "red flag" test.[562] This test has led to reliance upon a willful blindness doctrine in establishing the knowledge requirement, namely taking actions to avoid confirming a high likelihood of wrongdoing.[563] The standard of willful blindness expands the boundaries of liability and creates ambiguity regarding its boundaries.[564] However, the Second Circuit remanded the case to the district court, which retreated from this approach, considered the different standards, and apparently merged "willful blindness" into the other two statutorily specified scienters (actual knowledge via takedown notices or "red flags")[565] when granting YouTube's summary judgment motion.[566] Today, most courts interpret this element as actual knowledge of a specific infringement.[567] The safe haven of the DMCA does not apply to an intentional wrongdoer.[568] In cases where an intermediary intentionally violates the law, it can be held liable—even if it lacks actual knowledge of a specific infringement. In other words, mere general knowledge of infringements suffices for imposing liability.[569] Based on the inducement doctrine, the Supreme Court held in *MGM Studios Inc. v. Grokster* that the intermediary was liable for

---

https://www.bna.com/court-affirms-dmca-n12884908880/ [https://perma.cc/7TCK-LW5Q]. Section 512(c)(1)(A)(i) refers to actual knowledge and § 512(c)(1)(A)(ii) refers to awareness of facts or circumstances from which infringing activity is apparent. Many discussions revolve around the meaning of the apparent standard of knowledge. *See, e.g.*, *Viacom Int'l Inc.*, 718 F. Supp. 2d at 518.

561. *See Viacom Int'l Inc.*, 718 F. Supp. 2d at 527. The N.Y. court clung to the actual knowledge test. *See id.* at 520. The second circuit was satisfied with objective "red flag" situations and willful blindness, and returned the case to the district court. *See* Viacom Int'l, Inc. v. YouTube, Inc., 676 F.3d 19, 31, 35 (2d Cir. 2012). The case ended in a settlement. *See* Stempel, *supra* note 451.

562. *See Viacom Int'l Inc.*, 718 F. Supp. 2d at 520–21.

563. *See* Methaya Sirichit, *Catching the Conscience: An Analysis of the Knowledge Theory Under § 512(C)'s Safe Harbor & the Role of Willful Blindness in the Finding of Red Flags*, 23 ALBANY L. J. SCI. & TECH. 85, 86, 109 (2013).

564. *See* David Welkowitz, *Willfulness™*, 79 ALBANY L. REV. 509, 510, 524 (2016).

565. *See* Eric Goldman, *Viacom Loses Again*–Viacom v. YouTube (Apr. 19, 2013), https://blog.ericgoldman.org/archives/2013/04/viacom_loses_ag.htm [https://perma.cc/X4VW-9MDN].

566. *See* Viacom Int'l Inc. v. YouTube, Inc., 940 F. Supp. 2d 110, 112, 115, 117 (S.D.N.Y. 2013). The case ended in a settlement. *See* Stempel, *supra* note 451.

567. *See, e.g.*, UMG Recordings, Inc. v. Shelter Capital Partners LLC, 718 F.3d 1006, 1015 (9th Cir. 2013); Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146, 1172 (9th Cir. 2007); UMG Recordings, Inc. v. Veoh Networks, Inc., 665 F. Supp. 2d 1099, 1105 (C.D. Cal. 2009); Corbis Corp. v. Amazon.com, Inc., 351 F. Supp. 2d 1090, 1099 (W.D. Wash. 2004).

568. *See* Miquel Peguera, *Secondary Liability for Copyright Infringement in the Web 2.0 Environment: Some Reflections on Viacom v. YouTube*, 6 J. INT'L COM. L. & TECH. 18, 24 (2011).

569. *See* Kohler, *supra* note 547, at 495; Peguera, *supra* note 568, at 20.

distributing file-sharing software to users with the intent to promote copyright infringement.[570] Since the intermediary induced its users to infringe copyrights by taking "active steps" to encourage direct copyright infringement,[571] it did not have dual purpose.[572]

The degree of intent is also discussed in various court decisions. Some courts prefer a high standard of intent and focus on the intermediary's desire to cause the harmful consequences.[573] Others choose a lower standard of intent and focus on the potential of the intermediary's act to bring about harmful consequences.[574] Together, these judicial decisions hold that an actor may be liable for intentionally encouraging direct infringement if the actor knowingly took steps that are substantially certain to result in such direct infringement.[575]

Scholarly work suggests that it is better to require actual knowledge to impose contributory liability, and a higher standard of intent to hold intermediaries liable for inducement—rather than adopting lower standards of knowledge and intent.[576] Clearer standards of knowledge and intent will maintain the DMCA's safe harbor and reduce uncertainty.[577]

c. Connecting the Dots Between Mental Element and Outcome

Tort cases involving intermediary influences on speech do not distinguish between contributory liability and inducement.[578] Copyright liability doctrines can provide courts with a toolbox of liability standards to distinguish among different situations. This

---

570. *See* MGM Studios Inc. v. Grokster, Ltd., 545 U.S. 913, 916, 918, 938 (2005) ("[O]ne who distributes a device with the object of promoting its use to infringe copyright, as shown by clear expression or other affirmative steps taken to foster infringement, is liable for the resulting acts of infringement by third parties.").

571. *See id.* at 923–24, 938.

572. *See id.* at 943 (citing Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 442 (1984)). The intent to encourage infringement made this case different from *Sony Corp. of America v. Universal City Studios, Inc.*, where there was no encouragement for infringement and the US Supreme Court concluded that a video recording device was capable of significant noninfringing use. *See Grokster*, 545 U.S. at 923–24; *Sony*, 464 U.S. at 446 n.28.

573. *See, e.g.*, *Grokster*, 545 U.S. at 938–39.

574. *Compare id.* at 935, *with* Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146, 1162 (9th Cir. 2007). See the interpretation of *Grokster* in *Perfect 10, Inc.* and *Columbia Pictures Industries, Inc. See* Columbia Pictures Indus., Inc. v. Fung, 710 F.3d 1020, 1037 (9th Cir. 2013); *Perfect 10, Inc.*, 508 F.3d at 1171.

575. *See Perfect 10, Inc.*, 508 F.3d at 1171.

576. *See* Mark Sableman, *ISPs and Content Liability: The Original Internet Law Twist*, THOMPSON COBURN (July 9, 2013), https://www.thompsoncoburn.com/insights/blogs/internet-law-twists-turns/post/2013-07-09/isps-and-content-liability-the-original-internet-law-twist [https://perma.cc/Z7FA-FTTF].

577. *See* Sirichit, *supra* note 563, at 144, 150, 186, 189.

578. *See* Inwood Labs., Inc. v. Ives Labs., Inc., 456 U.S. 844, 854 (1982); Perfect 10, Inc. v. Visa Int'l Serv. Ass'n, 494 F.3d 788, 796 (9th Cir. 2007).

Subsection provides guidelines for adjusting this toolbox for nudges. It first differentiates contributory liability and inducement. Then, it clarifies the standards of knowledge and intent in the context of nudge torts.

This new framework proposes that courts should hold an intermediary responsible for users' defamatory content when it knowingly contributes to the distribution of users' defamatory speech. Courts should apply the actual knowledge test and not settle for other lesser forms of knowledge.[579] Yet, when an intermediary encourages speech torts with an intent to promote defamation, the inducement standard should apply.[580] Nudging users to generate defamatory content functions in a similar way to providing users with software that facilitates infringement or directly aids in infringement.[581] In such cases, there is a reason to argue that general knowledge of the act suffices to hold the intermediary responsible.[582] This interpretation might even allow the imposition of liability when the intermediary automatically and systematically pushes users to commit speech torts by using algorithms.[583] It assumes that the intermediary has general knowledge of the strong nudges it creates, and that it should avoid using "evil algorithms" in the first place.[584] This interpretation balances actual knowledge of specific illegal content (under contributory liability) and a subjective element of intent and general knowledge (under the inducement doctrine).[585]

These standards of liability allow courts to impose differential levels of liability. Consequently, liability could be imposed on intermediaries who nudge users to disseminate negative content, even

---

579. *See* Section III.A.1.a. On other forms of knowledge, see *Viacom International Inc. v. YouTube, Inc.*, 718 F. Supp. 2d 514, 519 (S.D.N.Y. 2010). Applying the actual knowledge test will limit the scope of liability and resolve the confusion left by the *YouTube* case. *See* Jacob Rogers, *YouTube v. Viacom: Second Circuit Ruling Leaves Open Possibility That YouTube Is Not Protected by Safe Harbor*, JOLT DIG. (Apr. 10, 2012), https://jolt.law.harvard.edu/digest/youtube-v-viacom [https://perma.cc/5YNT-U9RY].

580. *See* McDonald, *supra* note 152, at 262, 274. When analyzing the inducement doctrine in this Article's context, courts should apply the higher standard of intent as pronounced in *MGM Studios Inc. v. Grokster*. *See* MGM Studios Inc. v. Grokster, Ltd., 545 U.S. 913, 930, 936 (2005).

581. *See Grokster*, 545 U.S. at 923–24, 936; McDonald, *supra* note 152, at 262, 274.

582. The parallel developments in liability for inducement in copyright law support the conclusion that in cases of inducement to speech torts, general knowledge suffices. For expansion on the inducement standard in copyright, see Burk, *supra* note 544, at 440; Kohler, *supra* note 547, at 495 (similar to the decision in *Grokster*, 545 U.S. at 941).

583. *See Grokster*, 545 U.S. at 923–24, 936; McDonald, *supra* note 152, at 262, 274.

584. *See Grokster*, 545 U.S. at 923–24, 936; McDonald, *supra* note 152, at 262, 274. On the ability to influence ex ante by design choices, see generally SKOVER &COLLINS *supra* note 392, at 27 (giving the example of Apple Siri's limitations on the culpability of the system); Mulligan & Bamberger, *supra* note 427; Scherer, *supra* note 434. On policy directed algorithms, see Tene & Polonetsky, *supra* note 430.

585. *See Grokster*, 545 U.S. at 930, 936, 944 n.1.

if the platform has legitimate uses and the nudge does not aim to push users to specifically generate defamatory content.[586] However, in such cases, courts should require a higher threshold of knowledge to impose liability. The scope of liability will be limited, and the intermediaries would be able to avoid liability by removing defamatory content ex post.[587] In contrast, in extreme cases of evil nudges involving unsavory actors working in bad faith moderation, the inducement doctrine allows the application of a lower threshold.[588] Courts can hold intermediaries responsible for any defamatory speech on their platforms, even in the absence of actual knowledge of specific defamatory content and in spite of the removal of offensive content ex post.[589] The intermediary will be considered as a joint tortfeasor with the direct user under the inducement standard.[590] This standard will function as a substantial negative incentive to avoid evil nudges ex ante.

The proposed framework allocates substantial negative incentives and deters the worst actors from the beginning. In addition, the framework re-allocates relatively moderate negative incentives to nudges that may have legitimate purposes but can cause harm.

d. Differential Standards as a Bridge Between Deontological and Consequential Perspectives

The proposed framework articulates both deontological and consequential considerations. It is based on the nature of the intermediary's conduct—the nudge—and elements of fault (actual knowledge or intent), both of which reflect corrective justice. It also considers efficiency and aggregated welfare since the outcome is tailored to the type of nudge.

A strong nudge—which pushes users to commit speech torts, or avoid participation otherwise[591]—is different from an ambiguous nudge.[592] When the degree of a nudge is strong, the intermediary should be held responsible under the inducement doctrine. In such

---

586. *See id.* at 926, 937, 939.

587. This regime is similar to the notice-and-takedown regime of 17 U.S.C. §§ 512(c), (j)(3) (2012).

588. *See Grokster*, 545 U.S. at 936.

589. *See* Brian C. McManus, Note, *Rethinking Defamation Liability for Internet Service Providers*, 35 SUFFOLK U. L. REV. 647, 653 (2001).

590. *See Grokster*, 545 U.S. at 930. On the doctrine of joint tortfeasors, see Wright, *supra* note 411, at 1142.

591. *See* THALER & SUNSTEIN, *supra* note 1, at 6; McDonald, *supra* note 152, at 262, 274. One can refer to a strong nudge as an "exclusionary vibe." *See* STRAHILEVITZ, *supra* note 128, at 43.

592. This differentiation is influenced by the notion of "capable of substantial non-infringing uses" in *Sony Corp. of America v. Universal City Studios, Inc.*, 464 U.S. 417, 442 (1984).

cases, an evidentiary presumption of intent applies. The degree of a nudge also indicates a causal link between the act (nudge) and the harmful outcome (defamatory speech). When the degree of a nudge is weaker, courts should examine the case under contributory liability doctrines. In such cases, there is no presumption of intent and the casual link with the outcome is more fragile.[593] Therefore, in this latter context, actual knowledge of specific defamatory content is a prerequisite for liability to attach.[594] Actual knowledge bridges between the intermediary's action and the outcome. It also demonstrates a causal link between the nudge and the defamatory speech.

e. The Optimal Regime

Applying differential standards leads to nuanced and proportional liability. The proposed framework is superior to the overall immunity regime, which is overinclusive and does not deter bad faith moderation,[595] may foster irresponsibility,[596] and undermines victims' freedom of expression.[597]

Settling on a standard of inducement for intermediaries who are by no means "good Samaritans" and immunizing all the rest is also underinclusive. Under this regime, intermediaries who use an implicit nudge, such as gossip and complaint websites, avoid liability. Due to the vast influence of nudges on the flow of information,[598] complete exemption from liability for intermediaries in this gray area is undesirable. A single standard of contributory liability regime for all nudges is not optimal either, since websites with no legitimate aim would take down harmful content only upon specific knowledge and continue to proliferate.

Applying a combination of standards—inducement and contributory liability—is superior to a negligence regime, since negligence is open-ended and leads to uncertainty. Negligence may also lead to hindsight and outcome biases because the reasonableness of the

593. *See Grokster*, 545 U.S. at 932–33.

594. *See* McManus, *supra* note 589, at 651–52. In the context of third-party liability to defamation, courts should apply the actual knowledge standard. On different standards of knowledge, see *Viacom International Inc. v. YouTube, Inc.*, 718 F. Supp. 2d 514, 519 (S.D.N.Y. 2010).

595. *See* Grimmelmann, *supra* note 314, at 103–05, 107.

596. *See* Citron & Wittes, *supra* note 386, at 413. ("An overbroad reading of the CDA has given online platforms a free pass to ignore illegal activities, to deliberately repost illegal material, and to solicit unlawful activities . . . . Companies have too limited an incentive to insist on lawful conduct on their services beyond the narrow scope of their terms of service. . . . They have no accountability for destructive uses of their services, even when they encourage those uses.").

597. *See* CITRON, *supra* note 76, at 194.

598. On this gray area, see THALER & SUNSTEIN, *supra* note 1, at 8.

action is decided after the fact.[599] In contrast, the proposed framework focuses on the nudge itself or on the actual knowledge of specific defamatory speech. These elements are clearer relative to the negligence standard and can promote accuracy and proportionality in legal responsibility.

2. The Proposed Framework in Action

a. Focal Point Nudges and Liability

An intermediary who pushes users to generate speech torts, such as TheDirty.com,[600] generates an illegitimate forum per se.[601] On this platform, it is very difficult for a user to avoid committing speech torts. This focal point leads to a presumption of intent to promote defamation on the part of the platform. Due to the degree of the nudge, the intermediary should be subjected to a heavy burden of liability under the inducement standard. Courts should hold the intermediary responsible for every instance of defamatory speech on the platform, even without actual knowledge of the specific defamatory speech and despite removing the defamatory speech ex post. This regime will disincentivize the operation of these platforms in the first place.

Yet, an intermediary who designs a focal point for complaints and negative content, such as BadBusiness.com,[602] does not exclusively encourage speech torts. These platforms may have legitimate purposes, and the nudge is weaker in comparison to a nudge on an illegitimate forum. While its influence is less extensive, it may also exacerbate the risk for speech torts. Therefore, contributory liability should apply.

---

599. On hindsight and outcome biases, see Yoed Halbersberg & Ehud Guttel, *Behavioral Economics and Tort Law*, *in* THE OXFORD HANDBOOK OF BEHAVIORAL ECONOMICS AND LAW 1, 2–3, 10 (Eyal Zamir & Doron Teichman eds., 2014) ("[H]indsight bias . . . distorts people's ex post assessments of the ex ante probability and predictability of an event, given that this event has already happened . . . . The outcome bias is the tendency to perceive conduct that resulted in a bad outcome as more careless than the same conduct in cases where the bad outcome did not occur."); Baruch Fischhoff, *Hindsight ≠ Foresight: The Effect of Outcome Knowledge on Judgment Under Uncertainty*, 1 J. EXPERIMENTAL PSYCHOL. 288, 295 (1975).

600. *See* Citron & Wittes, *supra* note 386, at 402; Knibbs, *supra* note 9. Platforms devoted specifically to defamation or hate speech have no dual use and their goal is limited to distributing illegal content. *See* Citron & Wittes, *supra* note 386, at 402, 413. Thus, even according to *Sony Corp. of America*, the intermediary should be held responsible. *See* Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 456 (1984).

601. *See Sony Corp. of Am.*, 464 U.S. at 456. It should be noted that, in some cases, the court reached an opposite conclusion and decided that the name of the platform does not indicate intent to encourage illegal content. *See, e.g.*, Perfect 10, Inc. v. CCBill LLC, 488 F.3d 1102, 1114 (9th Cir. 2007); UMG Recordings, Inc. v. Veoh Networks, Inc., 665 F. Supp. 2d 1099, 1107, 1109, 1111 (C.D. Cal. 2009). However, in the context of speech torts, the name of the platform should be a sufficient reason for liability due to its far-reaching effects on the gravity of harm.

602. *See* Hy Cite Corp. v. Badbusinessbureau.com, L.L.C., 418 F. Supp. 2d 1142, 1145 (D. Ariz. 2005).

The nudge itself does not indicate a mental element of intent, and courts should hold intermediaries responsible only upon actual knowledge of a specific speech tort. In such instances, liability can be avoided by removing the speech ex post.

b. Channeling and Leading Nudges and Liability

Similar to focal points, courts should examine the domain of choice given to users in drop-down menus or filter mechanisms.[603] Some intermediaries do not leave users a broad spectrum of choice and push them to choose a defamatory option.[604] The Ninth Circuit's opinion in *Roommates.com* is based on this situation.[605] In such cases, the degree of the nudge indicates the element of intent and the effect of channeling and leading is substantial. Therefore, the intermediary should be subjected to the heavy burden of the inducement standard. This conclusion is reinforced when the intermediary uses data mining and artificial intelligence to create personalized nudges that match the characteristics of every user and specifically push users to illegitimate forums, or to be engaged in illegal activities. Unlike drop-down menus that allow users to choose among multiple options, explicit personal recommendations on illegal content channel the user to participate in illegal activities without a meaningful domain of choice between options. Instead, the intermediary chooses what recommendations users see. Therefore, these systems should not be considered "neutral tools."[606] Applying the standard of inducement is likely to cause the intermediary to amend the list of pre-made choices that lead to defamatory speech, or avoid recommending illegitimate content altogether.[607]

---

603. *See* Percival, *supra* note 278, at 171. Courts should examine the choice options given to users, how many of them are illegal, and whether specific choices are emphasized or preferred over others. *See id.*

604. *See id.*

605. *See* Appellants' Opening Brief at 27, Fair Hous. Council v. Roommates.com, LLC, 489 F.3d 921 (9th Cir. 2007) (No. 04-56916) ("Roommates tells users to select between A and B, and where both A and B are discriminatory. . . . By creating the two discriminatory choices and telling the user to select among them, Roommates plays a 'significant role' in the provision of the information at issue.").

606. This conclusion is not in line with *Dyroff v. Ultimate Software Group, Inc.*, No. 17-CV-05359-LB, 2017 WL 5665670, at *14 (N.D. Cal. Nov. 26, 2017). Yet, the law should differentiate between tools that allow users to choose and tools that strongly push users to engage in illegal content. The latter should not be considered neutral, and an intermediary that has general knowledge of illegal recommendations should bear liability. Recommending illegitimate forums and pushing users directly to it is different from just hosting illegal content. Therefore, an intermediary should avoid illegal recommendations in the first place. Indeed, this may result in less recommendations, or some degree of chilling on innovation. Yet, in such cases, a degree of chilling is worthwhile. *See id.*

607. *See* Wu, *supra* note 368, at 296, 300, 343.

When the intermediary allows a broad list of choices,[608] it does not push users to commit speech torts. The nudge is weaker and its influence is less substantial. However, designing imbalanced menus, which include defamatory options, may also exacerbate harm. Therefore, courts should apply the standard of contributory liability and hold the intermediary responsible if defamatory expressions are not removed ex post.

c. Encouragement Nudges and Liability

Courts should examine whether an intermediary's encouragement is concrete and specific.[609] A strong nudge specifically pushes users to commit speech torts or avoid participation. In such cases, responsiveness to the intermediary's explicit push would surely lead users to disseminate defamatory expressions. The nudge provides an indication of the intermediary's intent and a causal link to speech torts. Therefore, there are strong justifications to impose the heavy burden of liability by using the inducement standard.[610] However, when the encouragement is implicit and participation still leaves a choice not to commit speech tort, it can be interpreted in more than one way and has a weaker influence.[611] Therefore, the normative standard should be contributory liability.

When the intermediary uses a combination of multiple nudging strategies to exacerbate harm,[612] courts should also apply the inducement standard.[613] In such cases, even if each one of the nudge strategies leaves users a choice to avoid defamation, together they have a cumulative effect. The combination of nudges indicates intent and a

608. *See* GW Equity LLC v. Xcentric Ventures LLC, No. 3:07-CV-976-O, 2009 WL 62173, at *13 (N.D. Tex. Jan. 9, 2009); Whitney Info. Network, Inc. v. Xcentric Ventures, LLC, No. 204-CV-47-FTM-34SPC, 2008 WL 450095, at *10 (M.D. Fla. Feb. 15, 2008). In these cases, only some of the options channeled users to negative content. *See GW Equity*, 2009 WL 62173, at *13–14; *Whitney*, 2008 WL 450095, at *10. The court granted immunity due to the wide domain of choice and unconsciously used the criteria of the degree of nudge. *See GW Equity*, 2009 WL 62173, at *14; *Whitney*, 2008 WL 450095, at *10, 12.

609. *See* Doty, *supra* note 207, at 132.

610. The slogan of the hypothetic platform Harassthem.com is "Don't Get Mad, Get Even." See *Fair Housing Council v. Roommates.com, LLC*, 489 F.3d 921, 928 (9th Cir. 2007). This slogan explicitly encourages users to harm others as revenge. *See* FTC v. Accusearch Inc., 570 F.3d 1187, 1195, 1199, 1200 (10th Cir. 2009).

611. *See FTC*, 570 F.3d at 1200; THALER & SUNSTEIN, *supra* note 1, at 6. For example, using slogan such as "Keep it Juicy" encourages gossip but not necessarily defamation.

612. *See* THALER & SUNSTEIN, *supra* note 1, at 72, 248. For example, an intermediary may use a combination of focal points, channeling and leading, and encouragement strategies. *See supra* Section II.D.

613. The lesson on cumulative nudges and an inducement standard is learned from the Supreme Court's decision in *MGM Studios Inc. v. Grokster, Ltd.*, 545 U.S. 913, 936 (2005).

causal link between the nudges and user speech torts. This cumulative effect should be considered a strong nudge.

Additionally, courts should disincentivize intermediaries who nudge users to commit speech torts and directly profit from the commission of those torts. Nudging speech torts and charging fees for their removal[614] should also indicate the intermediary's intent and lead to liability under the inducement standard.[615]

In sum, the proposed framework allows courts to impose nuanced burdens of liability depending on the nudge and bridge between the deontological and consequential perspectives. It also strikes an optimal balance between free speech and reputation.[616] The combination of inducement and contributory liability allows different levels of deterrence depending on the degree of a nudge. Finally, they promote corrective justice,[617] efficiency,[618] and innovation.[619]

Table 2. Summary of the Guidelines for Differential Liability to Tort Nudges

Standard of Liability	Focal Points	Channeling and Leading	Encouragement
Inducement	▪ Strong nudges that push users to commit speech torts and constitute "illegitimate forums"	▪ Designing menus, filters, or tags that include only options	▪ An explicit encouragement to commit speech tort. This nudge pushes users to commit

614. *See* Glob. Royalties, Ltd. v. Xcentric Ventures, LLC, 544 F. Supp. 2d 929, 930 (D. Ariz. 2008); RIPOFF REPORT, *supra* note 201.

615. On these programs, see *MGM Studios Inc. v. Grokster, Ltd.*, 545 U.S. 913, 930 (2005); *Glob. Royalties*, 544 F. Supp. 2d at 930.

616. *See* Lavi, *supra* note 71, at 933. Differential standards of liability lead to proportional chilling effects and strike an optimal balance between various constitutional rights. *See supra* Section II.B.1. Such standards will not hinder the open market of ideas disproportionately. *See* Lavi, *supra* note 71, at 933.

617. *See supra* Section III.B.2.a. According to corrective justice considerations, when the influence of a nudge is low, the action itself (the nudge) does not fulfill the element of wrongful cause of harm. *See* Lavi, *supra* note 61, at 183. In such cases, a more profound mental element (actual knowledge) is required in order to impose liability. *See id.* at 183. In contrast, a strong nudge reflects the element of cause indicating intent. *See id.* at 184. Imposing liability for strong nudges on their own can be justified by corrective justice considerations. *See id.*

618. *See supra* Section III.B.2.b. Liability for strong nudges creates more benefits than costs. *See* Lavi, *supra* note 61, at 187. However, when the nudge is ambiguous, the costs of avoidance are higher. *See id.* at 186. In such cases, it would be inefficient to discourage the action itself; rather, a different standard of liability should be imposed—a contributory liability regime. *See id.*

619. *See* Kim, *supra* note 11, at 424. This Article's proposed framework does not impose liability on technology, but rather on action and the mental element that adds value to the technology. *See supra* Section III.B.2.c. Moreover, the framework avoids curbing technological development and can adapt to changing technologies. *See* Kim, *supra* note 11, at 421; *supra* Section III.B.2.c.

Standard of Liability	Focal Points	Channeling and Leading	Encouragement
	(e.g., The Dirty.com). ▪ Combining multiple strategies that promote libel. Even if each one of the strategies provides users with choice, together they have a cumulative effect that should be considered a strong nudge. ▪ Nudging defamatory content and charging fees for its removal. ▪ The *nature of the act* (the nudge) itself indicates the intent of the intermediary (subjective *mental element*) and provides a *causal link* between *the act* and *the outcome* of speech torts. This standard of liability will deter generation of evil nudges ex ante and lead to efficiency. It will also likely cause intermediaries to avoid creating illegitimate forums. ▪ Courts may find the intermediary responsible for defamation even if it removed it ex post.	that will necessarily lead to speech torts. ▪ When the intermediary only offers defamatory options in menus and pushes users to choose between participating and committing speech torts to avoiding participation (e.g., Roommates.com). ▪ Combining multiple strategies for influencing speech tort. ▪ Nudging defamatory content and charging fees for its removal. ▪ The *act* indicates the intent of the intermediary and a *causal link* to the *outcome*. The inducement standard of liability for extreme cases of channeling and leading functions as an ex ante negative incentive to design architecture that specifically channels users to disseminate speech torts.	speech torts or avoid participation, and favors the first option (e.g., TheDirty.com, "Submit dirt"; "Is Anyone Up"; "Pure evil"). ▪ Combining multiple strategies for influencing speech torts. ▪ Nudging defamatory content and charging fees for its removal. ▪ Applying the inducement standard of liability for explicit nudges functions as an ex ante negative incentive to explicitly encourage speech torts.
Contributory Liability	▪ An implicit nudge in the gray area (e.g., BadBusiness.com). ▪ This type of nudge promotes speech	▪ Designing menus, filters, or tags which are unbalanced and include more negative options than positive and	▪ An implicit nudge in the gray area (e.g., "keep it juicy"; "Don't let them get away with it").

Standard of Liability	Focal Points	Channeling and Leading	Encouragement
	torts but provides users with a broad choice other than disseminating illegal content. The degree of influence is lower relative to illegitimate forums. ▪ The *act* (nudge) does not indicate a *mental element* or a *causal link* to the *outcome* (speech tort) on its own. *Actual knowledge* to specific defamation will bridge the gap and allow courts to impose liability under a contributory liability standard. ▪ The scope of liability is limited and the intermediary can avoid it by removing the defamatory content ex post. Consequently, this regime does not discourage the creation of focal points, which can have legitimate purposes. The negative incentive focuses on the removal of defamatory content ex post.	neutral ones (e.g., RipoffReport.com). ▪ Emphasizing the negative options in menus. ▪ The intermediary does not specifically nudge users to disseminate speech torts, however by designing imbalanced biased options, he channels them and indirectly promotes speech torts. ▪ Applying the standard of contributory liability allows courts to impose liability only when the intermediary has actual knowledge of a specific defamatory speech on his platform and avoids ex post removal. ▪ The *actual knowledge* bridges the gap between the act and the outcome. The scope of liability is limited and will not chill platforms' design disproportionately.	▪ This type of nudge promotes speech torts but provides users with a broad choice, other than disseminating defamation. ▪ The degree of influence is lower relative to explicit nudges. ▪ Applying the standard of contributory liability allows courts to impose liability only when the intermediary has actual knowledge of a specific defamatory speech on his platform and avoids ex post removal. ▪ The *actual knowledge* bridges the gap between the act and the outcome. The scope of liability is limited and will not chill encouragement and efficient moderation in general.

B. The Proposed Framework and the Law Bridging the Gaps

The current law provides immunity for intermediaries,[620] such that they are not treated as publishers for material they did not author

620. *See* Communications Decency Act, 47 U.S.C. § 230 (2012); *supra* Section III.A.1.

or develop.[621] Courts usually interpret this immunity broadly.[622] However, this overall immunity scheme was constructed when the web was at its genesis.[623] As technologies advance and the web becomes more prevalent, the increased potential for online torts leads to a substantial increase in the gravity of harm.[624] Therefore, it is time to challenge the immunity regime and refine it.[625]

A large body of scholarly work suggests one way to narrow immunity is by amending section 230.[626] Some suggest that immunity should not apply to intermediaries that materially contribute to illegal or tortious content[627] and propose ways for Congress to revise section 230 to withdraw protections from such intermediaries.[628] A broader

---

621. *See* § 230; *supra* Section III.A.1. There are some exceptions to the immunity. *See* § 230(e)(1)–(5). This Article focuses on defamation in civil law, but it should be noted that immunity is limited to civil claims and does not apply to cases that are based on federal criminal law. *See* § 230(e)(1). In addition, the Senate recently passed a bill that holds online intermediaries accountable for third-party content that encourages sex trafficking. *See* Allow States and Victims to Fight Online Sex Trafficking Act, H.R. 1865, 115th Cong. (2018); Zeynep Ulku Kahveci, *Allow States and Victims to Fight Online Sex Trafficking Act (FOSTA): Senate Passes Bill Making Online Platforms Liable for Third-Party Content Enabling Illegal Sex-Trafficking*, JOLT DIG. (Apr. 4, 2018), https://jolt.law.harvard.edu/digest/allow-states-and-victims-to-fight-online-sex-trafficking-act-fosta-senate-passes-bill-making-online-platforms-liable-for-third-party-content-enabling-illegal-sex-trafficking [https://perma.cc/6BZS-VCZH]; Eric Goldman, *'Worst of Both Worlds' FOSTA Signed Into Law, Completing Section 230's Evisceration*, TECH. & MARKETING L. BLOG (Apr. 11, 2018), https://blog.ericgoldman.org/archives/2018/04/worst-of-both-worlds-fosta-signed-into-law-completing-section-230s-evisceration.htm [https://perma.cc/65JY-FSPX]. For another perspective, see Mary Graw Leary, *The Indecency and Injustice of Section 230 of the Communications Decency Act*, 41 HARV. J.L. & PUB. POL'Y 553, 620 (2018).

622. *See supra* Section III.A.1.

623. *See* Leary, *supra* note 621, at 574 ("In 1997, when cases first percolated through the court system, the Internet was in its infancy.").

624. *See* Citron & Wittes, *supra* note 386, at 411–12; OLIVIER SYLVAIN, KNIGHT FIRST AMENDMENT INST., COLUMBIA UNIV., DISCRIMINATORY DESIGNS ON USER DATA 12 (2018) ("[T]hese developments undermine any notion that online intermediaries deserve immunity because they are mere conduits for, or passive publishers of, their users' expression."). Online intermediaries pervasively shape, study, and exploit communicative acts on their services and with greater power comes greater potential for harm. *See* GILLESPIE, *supra* note 55, at 43 ("[T]he moment that a platform begins to select some content over others, based not on a judgment of relevance to a search query but in the spirit of enhancing the value of the experience and keeping users on the site, it [becomes] a hybrid [of conduit and media].").

625. *See* Sylvain, *supra* note 246, at 208.

626. *See, e.g.*, DANIELLE KEATS CITRON, KNIGHT FIRST AMENDMENT INST., COLUMBIA UNIV., SECTION 230'S CHALLENGE TO CIVIL RIGHTS AND CIVIL LIBERTIES 6–7 (2018); Citron & Wittes, *supra* note 386, at 418 ("Platforms should enjoy immunity from liability if they could show that their response to unlawful uses of their services was reasonable."); Sylvain, *supra* note 246, at 214 (urging Congress to maintain the immunity but to create an explicit exception from the safe harbor for civil rights violations).

627. *See* Citron & Wittes, *supra* note 386, at 419 ("A modest alternative to a sweeping elimination of the immunity for state law would be to eliminate the immunity for the worst actors. . . . [S]ites that encourage destructive online abuse or that know they are principally used for that purpose should not enjoy immunity from liability.").

628. *See id.*; Cecil, *supra* note 285, at 2549; Goldnick, *supra* note 493, at 626–27. Some scholars even suggest revising § 230 to include a general notice-and-take-down provisions. *See*

approach is revising section 230 and conditioning intermediary exemption from liability in taking reasonable steps to prevent or address unlawful uses of its services.[629]

However, courts—without legislative changes—can set the proper boundaries of immunity.[630] Applying the proposed guidelines allows courts flexibility in accommodating the dynamic online environment.[631] According to a proper reading of section 230, intermediaries that nudge speech torts are "responsible" at least "in part" for creating or developing defamatory content and should not enjoy the immunity.[632]

Courts can broadly interpret the decision in *Roommates.com* to narrow section 230's immunity.[633] Due to the increasing influence of nudges in the digital age, as well as the substantial harm they cause, this provides the proper solution.

C. Addressing Objections to the Proposed Framework

Several objections to this framework can be anticipated—thus, some wrinkles must be ironed out. The first objection is directed at the very idea of acknowledging liability for evil nudges. One may argue that even a strong explicit nudge that pushes users to commit speech torts always leaves users the option of ignoring the nudge or avoiding participation.[634] As some courts have stated, users are not forced to

---

Vanessa S. Browne-Barbour, *Losing Their License to Libel: Revisiting § 230 Immunity*, 30 BERKELEY TECH. L.J. 1505, 1554 (2015).

629. *See* Danielle Keats Citron, *Sexual Privacy*, YALE L.J. (forthcoming 2019) ("Modest adjustments to Section 230 could maintain a robust culture of free speech online without extending the safe harbor to bad actors or, more broadly, to platforms that do not respond to illegality in a reasonable manner."); Citron & Wittes, *supra* note 386, at 419; Citron & Wittes, *supra* note 314, at 471.

630. *See* Tremble, *supra* note 250, at 867. For suggestions on judicially narrowing immunity without additional amendments, see *id.*; Sylvain, *supra* note 246, at 214 (suggesting an even narrower interpretation of the immunity by courts since some intermediaries are not passive conduits). The interpretative route is also preferred by Citron and Wittes. *See* Citron & Wittes, *supra* note 386, at 418 ("If the courts decline to move § 230 in this direction, Congress should consider statutory changes.").

631. *See* Tremble, *supra* note 250, at 867.

632. *See* 47 U.S.C. § 230(c), (f)(3) (2012); Franklin, *supra* note 243, at 1334.

633. *See* FTC v. Accusearch Inc., 570 F.3d 1187, 1198 (10th Cir. 2009); Vision Sec., LLC v. Xcentric Ventures, LLC, No. 2:13-cv-00926-CW-BCW, 2015 WL 12780892, at *2 (D. Utah Aug. 27, 2015) (concluding that a service provider is not neutral if it "specifically encourages development of what is offensive about the content") (citing Fair Hous. Council v. Roommates.com, LLC, 521 F.3d 1157 (9th Cir. 2008))).

634. *See, e.g.*, GW Equity LLC v. Xcentric Ventures LLC, No. 3:07-CV-976-O, 2009 WL 62173, at *5 (N.D. Tex. Jan. 9, 2009).

generate illegal content,[635] leaving no sufficient justification to impose intermediary liability.

Indeed, individuals are not forced to generate speech torts. The possibility that a user can avoid generating a speech tort, despite the nudge, raises complex questions regarding the limitations of liability for nudges. However, in the context described in this Article, online intermediaries use sophisticated technologies and strategies to exert hyper-control and influence over their platforms. These tools can efficiently influence users to commit speech torts and lessen the control over their own decisions to publish content.[636] As demonstrated, intermediaries utilize users' cognitive biases and influence the flow of information.[637] Due to intermediaries' centralized power, their nudges are far more influential than those nudges generated by an average individual user. Furthermore, the internet forms an exceptional context.[638] On the internet, nudges have an extensive effect on social networks and the flow of information.[639] Consequently, the push is so strong that it may be very difficult for a user to avoid speech torts.[640] In this unique setting, intermediary nudges substantively increase the likelihood and gravity of harm.[641] Therefore, there are strong justifications to hold an online intermediary responsible for evil nudges.

The second objection is the potential for overdeterrence imposed by the proposed framework relative to the current immunity regime. The proposed guidelines do not provide a precise formulation for proscribed conduct.[642] They require weighing the degree of a nudge, and may lead to uncertainty.[643] The answer to the question of which type of nudge will exclude an intermediary from immunity remains inconclusive. Outside the scope of the worst actors that should be subjected to an inducement standard, the scope of liability is vague. Due to this ambiguity, more motions to dismiss will be denied—

635. *See id.* at *5–6; Whitney Info. Network, Inc. v. Xcentric Ventures, LLC, No. 204-CV-47-FTM-34SPC, 2008 WL 450095, at *11 (M.D. Fla. Feb. 15, 2008).

636. *See generally* Wang et al., *supra* note 378 (addressing the ability of nudges to influence self-control).

637. *See generally id.* Intermediaries use insights on network structures and technologies, such as big data and artificial intelligence, to efficiently nudge. *See supra* Section II.D.

638. On this perspective, see LIPTON, *supra* note 277, at 4.

639. See RAINIE & WELLMAN, *supra* note 51, at 285 (describing the news ecology that digital networks created). In the context of nudges, the technological ecosystem makes it easier to nudge—resulting in more influential nudges at the network level. *See id.* In addition, new technologies like big data and AI allow intermediaries to create nudges with accuracy and manipulate users. *See* Yeung, *supra* note 115, at 15.

640. *See* Yeung, *supra* note 115, at 8. On the influence of nudges in the age of big data, see *id.*

641. *See* Levi, *supra* note 19 (manuscript at 26); *supra* Section II.D.

642. *See supra* Section IV.A.2.

643. *See supra* Section IV.A.1.c.

allowing lawsuits to proceed from preliminary stages. Accordingly, this will increase administrative costs. Intermediaries who are neutral to tortious content may also act defensively and remove any content in response to complaints, even if it is not defamatory. This results in censorship and stifles the development of innovative platforms.

Indeed, the proposed guidelines reduce certainty relative to the overall immunity or other rule-based formulations of liability. However, balancing the overall costs and benefits against the alternatives leads to the conclusion that relative ambiguity is a worthwhile price. The alternative of a rule-based formulation for liability may entail more certainty, but will lead to distortions and less accuracy by being both over and underinclusive.[644]

A nuanced liability creates more benefits than shortcomings. The framework outlines differential standards, which include different elements and thresholds of liability that allow fitting proportional negative incentives to different degrees of nudges. This regime promotes efficiency more than other proposals reviewed in scholarly work.[645] In addition, today, more than a third of claims already survive a section 230 immunity defense.[646] The proposed guidelines structure judicial discretion, assist courts in applying open-ended standards, and adjust intermediary liability in torts for nudges. By structuring judicial discretion, courts are likely to reach more consistent, just, and efficient outcomes relative to the inconsistency reflected in the case law today. Certainty and consistency will grow over time as precedents applying the proposed guidelines accumulate.

The third objection is directed at nonsalient or nontransparent nudges. This objection argues that the guidelines do not fit them because it is difficult to recognize this type of nudge and its influences. Thus, with regard to these types of nudges, the guidelines result in underdeterrence.[647] For example, channeling and leading nudges are non-salient and their aim to influence speech torts is not obvious to internet users.[648] In contrast to direct persuasions to commit speech

644. *See* Lavi, *supra* note 71, at 859; *supra* Section III.C. For instance, an overall immunity regime will not disincentivize intermediaries to design illegal forums. *See* Lavi, *supra* note 71, at 885–86. Consequently, they will continue to use various nudging strategies and push users to publish and disseminate speech torts. *See id.* at 886. An overall "notice-and-takedown" safe haven does not always lead to an optimal level of deterrence and can be over- or underinclusive. *See id.* at 887. Thus, it will not bring to an optimal level of deterrence when the intermediaries explicitly nudge users to generate speech torts without leaving room to make their choice. *See id.*

645. *See supra* Section IV.A.2.

646. On this inconsistency, see *supra* Section II.A.1. According to an empirical study, more than a third of the claims survive a § 230 defense. *See* Ardia, *supra* note 240, at 392; Kosseff, *supra* note 282, at 20.

647. *See* Sunstein, *supra* note 24 (manuscript at 35).

648. *See supra* Section II.D.2.

torts, imbalanced options in drop-down menus apply to non-deliberative thinking and influence decision making subconsciously.

Additionally, with some encouragement nudges, the victim of speech torts is not aware of the nudge and its influences on the social network's context.[649] Today, intermediaries can nudge only some users, for instance, in the "influential" hubs in the social network.[650] They can also personalize messages, appeal to specific users in private messages, and encourage users to disseminate defamatory content.[651] As described above,[652] many intermediaries collect data on users and use artificial intelligence and complex algorithms to target their nudges more efficiently and exacerbate harm.[653]

When an intermediary applies nontransparent strategies, a victim of speech torts may be unaware of the intermediary's contributory liability or inducement.[654] Consequently, that victim will not be able to prove the intermediary's liability. The result of imposing liability on nontransparent nudges will be underdeterrence and inefficiency. This argument is valid; however, it focuses on specific situations of nontransparent nudges such as channeling and leading that are nonsalient even to the user and some of the encouragement that is nontransparent to third parties. Thus, it does not undermine the proposed guidelines. Market forces and complementary suggestions can mitigate the problem of underdeterrence and narrow the gap. For instance, a user may reveal the goal behind a particular nontransparent channeling and leading nudge and publicize it.[655] Alternatively, an "influential" user in the social network who has been subjected to a

649. *See supra* Section II.D.3.

650. *See* Aral & Walker, *supra* note 69, at 337.

651. *See* TUROW, *supra* note 60, at 139. On intermediary abilities locating the central hubs in the platform and conveying specific messages to them, see *id.*; Aral & Walker, *supra* note 69, at 337; Danielle Keats Citron & Neil M. Richards, *Four Principles for Digital Expression (You Won't Believe #3!)*, 95 WASH. U. L. REV. 1353, 1362 (2018) ("[O]pportunities are neither limitless nor uniform.").

652. *See supra* Section II.D.

653. *See* Levi, *supra* note 19 (manuscript at 26). In an experiment, Facebook showed some users fewer posts containing emotional language. *See* Grimmelmann, *supra* note 66, at 222. Facebook discovered that users who saw fewer positive posts used more negative words. *See id.* Facebook generated a nontransparent encouragement nudge to disseminate negative content. *See id.* For another recent example of nontransparent nudges, see *Dyroff v. Ultimate Software Group, Inc.*, No. 17-cv-05359-LB, 2017 WL 5665670, at *3 (N.D. Cal. Nov. 26, 2017). On opaque processes of algorithms, see PASQUALE, *supra* note 395, at 6. On the use of big data and artificial intelligence by intermediaries, see Balkin, *supra* note 54, at 1184.

654. *See supra* Section II.D.2.

655. *See* Shmuel Becher & Tal Zarsky, *Seduction by Disclosure: Comment on Seduction by Contract*, 9 JERUSALEM REV. LEGAL STUD. 72, 76 (2013).

poorly timed nudge may simply perceive this message as a nuisance.[656] In response, he might make the public aware of this practice, thus, bridging the information gap.[657] Due to the intermediary's concern for its reputation, it may ex ante avoid this strategy. In addition, regulators can call upon, or even fund, independent researchers specifically to analyze digital practices and attempt to uncover biased algorithms and manipulative practices of intermediaries' evil nudges.[658] These solutions have the potential to mitigate this problem. Nevertheless, they would reveal only some of the cases of nontransparent manipulative nudges to the public.

Yet, the guidelines do not preclude complementary, related legal adjustments that may mitigate this problem. One complementary solution may be imposing transparency obligations on intermediaries to disclose their nudging policy.[659] Disclosure of nudging strategies may increase the awareness of prospective targets to the intermediary's contribution to their potential harm—thus aiding actual victims in their attempts to satisfy their burdens of proof in courts. However, even if the intermediary complies with disclosure requirements, this solution appears to be insufficient because users often do not read or comprehend disclosures.[660]

---

656. *See* FOGG, *supra* note 15, at 43. On the importance of timing, see ALESSANDRO ACQUISTI ET AL., TIMING IS EVERYTHING? THE EFFECTS OF TIMING AND PLACEMENT OF ONLINE PRIVACY INDICATORS 319 (2009).

657. *See* Becher & Zarsky, *supra* note 655, at 76. On flows of information among internet users, see *id.*

658. *See* Ryan Calo & Alex Rosenblat, *The Taking Economy: Uber, Information, and Power*, 117 COLUM. L. REV. 1623, 1684 (2017) (focusing on nontransparent, manipulative practices in a related context of sharing economy platforms and suggesting that third party independent research can reveal some of these manipulative practices—thus having the potential for mitigating the problem of nontransparent evil nudges of intermediaries); Niva Elkin-Koren & Maayan Perel, *Algorithmic Governance by Online Intermediaries*, *in* OXFORD HANDBOOK OF INTERNATIONAL ECONOMIC GOVERNANCE AND MARKET REGULATION (forthcoming 2018) (manuscript at 16–17) (focusing on a related context of copyright algorithmic enforcement, which is committed without transparency, and proposing that private initiatives committed to protecting online free speech can retrieve information on improper practices of intermediaries and increase awareness among policy makers, the press, and the public for online violations); Maayan Perel & Niva Elkin-Koren, *Black Box Tinkering: Beyond Disclosure in Algorithmic Enforcement*, 69 FLA. L. REV. 181, 181 (2017) (proposing that the public can tinker the algorithmic black box and reveal improper algorithmic enforcement).

659. *See* Sunstein, *supra* note 24 (manuscript at 35). On virtues of transparency in related contexts, see Maayan Perel & Niva Elkin-Koren, *Accountability in Algorithmic Copyright Enforcement*, 19 STAN. TECH. L. REV. 473, 478 (2016) (focusing on a related context and suggesting transparency and public oversight to mitigate the problem in the context of algorithmic copyright enforcement); Tal Z. Zarsky, *Transparent Predictions*, 2013 U. ILL. L. REV. 1503, 1540 (2013).

660. On this point, see SUNSTEIN, *supra* note 25, at 150; Sunstein, *supra* note 24 (manuscript at 37); *supra* Section II.A. Additionally, transparency comes at a price and does not always lead to efficiency and fairness. *See* Tal Zarsky, *The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making*, 41 SCI. TECH. & HUM. VALUES 118, 122 (2016).

A better complementary legal policy is bridging the deterrence gap by adjusting compensation in these situations. Scholarly work aimed at solving tax evasion proposes a similar solution.[661] In the context of this Article, courts can impose higher compensation for nontransparent nudges, which are rarely discovered. Adjusting compensation to the probability of enforcement will increase the expected compensation for nontransparent nudges and disincentivize these strategies.[662] In addition, when courts find the intermediary liable for inducement in nontransparent ways, there is a justification for awarding punitive damages.[663] Such cases might also lead to governmental investigation and penalties for unfair or deceptive acts, designs, or practices under the FTC Act.[664] Nuanced compensation and penalties that are sensitive to nontransparent strategies and account for the probability of enforcement may narrow the deterrence gap.

V. CONCLUSION

This Article is the third in a series of scholarship that advances a context-based theory of liability to speech torts.[665] This Article aspires to take initial steps to address online intermediaries' contributory liability in cases of speech torts. It demonstrates that intermediaries can and do influence social contexts. They use different nudging strategies—pushing users to generate specific types of content, while influencing social dynamics and affecting the flow of information. It further shows that nudging strategies exacerbate harm. Therefore, overall immunity should not apply to intermediaries that push users to disseminate defamatory speech. Following this conclusion, this Article applies multidisciplinary insights to legal policy and offers an innovative, theoretical, and practical framework for regulating

661. *See* Alex Raskolnikov, *Crime and Punishment in Taxation: Deceit, Deterrence, and the Self-Adjusting Penalty*, 106 COLUM. L. REV. 569, 599 (2006).

662. *See id.* at 571. Ex ante risk management takes into account the probability for paying compensation and the amount. *See id.* at 602–03. A larger amount of compensation for nontransparent nudges balances the low probability for discovery and thus lead to more effective disincentives to use this type of nudge. *See id.*

663. On punitive damages, see RESTATEMENT (SECOND) OF TORTS § 908(1)–(2) (AM. LAW INST. 1979); Volker Behr, *Punitive Damages in America and German Law – Tendencies towards Approximation of Apparently Irreconcilable Concepts*, 78 CHI.-KENT L. REV. 105, 105 (2003); Michael L. Rustad & Thomas H. Koenig, *Taming the Tort Monster: The American Civil Justice System as a Battleground of Social Theory*, 68 BROOK. L. REV. 1, 60 (2002).

664. *See* Federal Trade Commission (FTC) Act of 1914 § 5, 15 U.S.C. § 45(m)(1)(a) (2018). Intermediaries' practices can be viewed as matters of consumer protection, privacy, data security, and technology policy.

665. *See* Lavi, *supra* note 71, at 855; Lavi, *supra* note 61, at 149. The first part focused on hosts' indirect liability. *See* Lavi, *supra* note 71, at 859. The second focused on intermediaries' direct liability. *See* Lavi, *supra* note 61, at 147–48.

intermediary nudges. Drawing from intermediary liability in copyright infringement, this Article proposes guidelines that apply differential standards of liability.

The guidelines outline different negative incentives depending on the degree of influence. Each aims to structure judicial discretion and assist courts in accommodating just and efficient policy. The guidelines direct courts to more systematic and consistent decisions and allow intermediaries, which are repeat players in court, to make ex ante predictions about the scope of their liability, which can lead to efficient risk management.

This framework also has broader influences beyond the scope of this Article. Currently, the law does not limit the influence of intermediaries on users' content as long as the intermediaries do not violate the law.[666] Intermediaries are free to influence online information and promote the generation and dissemination of content that is in-line with their ideological or commercial goals.[667] However, changes to intermediaries' incentives in the context of speech torts may have indirect effects on other contexts as well. Outlining specific procedures that apply only to nudging defamation may be complex since the line between defamatory speech and other types of speech is not always clear.[668] An intermediary that aims to run efficient risk-management and reduce its exposure to liability may avoid generating strong nudges and leave users with broader choices in general. The proposed guidelines can also extend to nudges that strongly push users towards arguably immoral or unethical behavior.[669] This may have an indirect effect in restraining undesirable nudges and incentivizing intermediaries to engage in fairer practices in general.

This Article constitutes the first sustained examination of the role of evil nudges in tort law. However, it is not the last word on this topic. Looking ahead, it inspires further discussions on intermediary liability in related contexts and liability for nudges in general. It raises intriguing legal questions on the scope of intermediary liability for

---

666. *See* Ardia, *supra* note 240, at 377.

667. *See* Calo, *supra* note 109, at 1001. On utilizing cognitive biases in online markets, see *id.*

668. *See* Lavi, supra note 61, at 178. For example, courts can decide that an expression benefits from defamation law defenses. *See id.* at 178–79. In addition, the line between a defamatory speech and other types of speech, such as privacy infringing speech and even criminal offences, is blurred online and different types of speech may overlap. *See* Anita Bernstein, *Real Remedies for Virtual Injuries*, 90 N.C. L. REV. 1457, 1464, 1468 (2012).

669. For example, Ashley Madison is a platform with a focal point on extramarital affairs that sports the slogan, "Life is short. Have an Affair." ASHLEY MADISON, https://www.ashleymadison.com/ [https://perma.cc/B2ZU-ENRW] (last visited Oct. 3, 2018).

creating focal points for nudging terror and incitement.[670] It also raises ethical and legal questions regarding the scope of intermediary liability for nudges that influence users to generate positive content. This type of nudge may, for instance, lead to glorifying specific products and mislead third parties about their market value. Should the law regulate these practices? When should the law consider nudges as manipulation?[671] How should the law react to the use of nontransparent nudges that push voters to vote for a specific candidate[672] or disseminate positive fake stories, which potentially influence elections, as exemplified by the Facebook-Cambridge Analytica scandal?[673]

Should the law hold intermediaries liable for nudges when they manipulate individuals and hinder their own self-interests, as opposed to those of third parties? Are lessons from the online experience transferable to other contexts of liability for nudges offline? What lessons should be learned for nudges at the age of the IoT that allows far more manipulative influences?[674] These are some challenges that should be discussed in future research.

---

670. *See* Fields v. Twitter, Inc., 881 F.3d 739, 743–44, 750 (9th Cir. 2018); Fields v. Twitter, Inc., 200 F. Supp. 3d 964, 968, 976 (N.D. Cal. 2016). This Article may allow courts to better interpret the scope of the Antiterrorism Act (ATA). *See* Antiterrorism Act (ATA), 18 U.S.C. §§ 2331, 2333 (2018).

671. This question is controversial. *See* Hansen & Jespersen, *supra* note 32, at 3; Wilkinson, *supra* note 378, at 342.

672. *See, e.g.*, Samuel, *supra* note 19.

673. *See* SCHNEIER, *supra* note 69, at 84; Zittrain, *supra* note 69, at 335, 336; Samuel, *supra* note 19; Tufekci, *supra* note 124. Some believe that Facebook used an algorithm to promote fake content in favor of then presidential candidate Donald Trump, and thus influenced the election results. *See id.*

674. *See* HILDEBRANDT, *supra* note 117, at 41 (referring to the elimination of the dichotomy between online and offline as the "onlife world"); SILVERMAN, *supra* note 15, at 300, 305; TUROW, *supra* note 122, at 19 (explaining that, in the future, in-store surveillance will be much more extensive and will build monitoring into people's routine activities); Michal S. Gal & Niva Elkin-Koren, *Algorithmic Consumers*, 30 HARV. J.L. & TECH. 1, 10–17 (2017) (explaining the potential of the IoT to improve customers' lives and addressing the objections and limitations of an algorithm that will improve customers shopping decisions and even decide on their behalf).