COMMUNITY VITALITY AS A THEORY OF GOVERNANCE FOR ONLINE INTERACTION

Farzaneh Badiei, Tracey Meares & Tom Tyler

OVERVIEW

Governance of platforms for online interaction has targeted primarily what users themselves put up online. That is, platform governance mechanisms typically focus on managing problematic content ranging from nudity to hate speech, something that platforms call "content moderation."[1] A review of both rules and strategies to enforce them reveals that moderation is focused on identifying and punishing bad behavior.[2] We think this is a mistake for at least two reasons. First, framing the issue of online content moderation primarily as an effort to find and suppress undesirable actions as opposed to focusing on strategies to encourage users to voluntarily internalize rules and engage in "good" behavior replicates the mistakes the criminal justice system has made in managing behavior identified as criminal "in the real world." Second, and perhaps more important, focusing on identifying and punishing bad behavior prioritizes elimination of bad behavior over the creation of a framework that facilitates healthful interaction. Such healthy interaction discourages the emergence of the bad behavior in the first place. We argue that just as in the real world it is better to facilitate and encourage healthful community interaction to avoid crime, platforms should engage in the project of creating infrastructures to encourage strong, healthful communities online.

---

[1] Robyn Caplan, *Data & Society, Content or Context Moderation? Artisanal, Community-Reliant, and Industrial Approaches* (2018), https://datasociety.net/library/content-or-context-moderation/.

[2] Tom Tyler et al., *Social Media Governance: Can Social Media Companies Motivate Voluntary Rule Following Behavior Among Their Users?*, 17 J. EXPERIMENTAL CRIMINOLOGY 109 (2019).

These communities are more likely to be self-regulating communities that need less external policing. In this article, we discuss how to import these ideas into the governance structure of platforms for online interactions.

We rely on our work with respect to the operation of the criminal justice system in the real world to demonstrate that organizing governance structures around the social psychology of procedural justice can produce positive results regarding voluntary compliance with rules and laws. People respond positively to procedural justice in the criminal justice system, in contrast to deterrence approaches premised upon the notion that people comply with rules and laws because they fear the consequences of failing to do so. Procedural justice strategies treat individuals as engaged agents who should have a part to play in the overall fair functioning of the system. Our experience working with platforms demonstrates that many rely on deterrence-based "get-tough" strategies to achieve compliance, and there is little reason to believe that such approaches work any better for social media than for criminal justice. Using procedural justice strategies to shape people's behavior online might prove useful.

Our larger point is that creating vital communities should be a key goal of governance as a general matter, whether on- or offline. We can divide online vital communities into two groups: those that engage in constructive and positive interactions on online platforms and those that use the opportunities provided by the platforms to manage their offline issues. An example of the former is the ongoing discussions that happen in Reddit or Facebook groups about race in the United States, stimulated by Black Lives Matter. An example of the latter is the efforts of online community groups to manage COVID-related problems in their communities—for example, Nextdoor's groups that offer help to the elderly and others in need

during the pandemic.[3] In both cases, the goal should be to leverage the possibilities of online platform communication to enhance community vitality and well-being. This involves lowering the level of negative or divisive online interaction and raising the constructive and problem-solving communication.

Building on theory and research demonstrating that people care more about the procedural fairness with which decision-makers treat them than the outcomes themselves, we explain how procedurally just treatment can encourage online community members to voluntarily follow platform rules and work constructively with each other to solve problems.

The first goal of this approach is to build self-regulatory models of content moderation. To do so, it is important to create commitment to rule-following that involves users' sense of obligation rather than their concerns about punitive measures, like having their posts or accounts blocked or permanently banned. To the degree that this model is effective, it is not necessary for platforms to try to identify wrongdoing. People more willingly follow the rules when they self-moderate than when coerced to take certain actions.

This approach has a second goal of building community vitality. The core of our argument is that the absence of harm is not the same thing as the presence of vitality. Community vitality is present when there are high levels of economic prosperity, social capital and well-being.

The goal of the suppression of harmful content may be a necessary beginning, but it is important to ask whether the strategies

---

[3] *Using Nextdoor to support your neighborhood during the COVID-19 pandemic,* NEXTDOOR, https://help.nextdoor.com/s/article/Using-Nextdoor-to-support-your-neighborhood-during-this-crisis? (last visited Feb. 15, 2021).

being used contribute to a long-term goal of building a vital community. If social media platforms adhere to procedural justice in their design, in their moderation efforts, and in their decision-making processes around the structuring of online groups, we believe that they can enhance community vitality and cooperation among online users. Those users work more constructively together, build shared identification and solidarity, and reach consensus approaches about how to address the issues that concern them.

INTRODUCTION

It was only in 2018 that Facebook first made public the guidelines that its moderators used to enforce its community standards.[4] This is not to say there had never been any rules. For twelve years, Facebook had prohibited publishing many types of objectionable content on its platform.[5] Despite the existence of these rules, however, users technically had no idea what the rules were or how Facebook enforced them unless they happened to violate them. When users did violate a Facebook content moderation rule, they were punished by being banned from using the platform for a particular period of time. For many years, Facebook unilaterally took content down, and there was no opportunity for the users to engage with the company about the consequence of a violation. Facebook's historical approach created a dynamic that remains a pillar of the relationship between social media platforms and their users: the platform imposes and enforces rules and the users obey.[6]

---

[4] Monika Bickert, *Publishing Our Internal Enforcement Guidelines and Expanding Our Appeals Process*, FACEBOOK (Apr. 24, 2018), https://about.fb.com/news/2018/04/comprehensive-community-standards/.

[5] Facebook, as early as 2006, had rules that governed the users and their content. *Member Content Posted on the Site,* FACEBOOK (Jan. 11, 2006), https://web.archive.org/web/20060118020625/https://www.facebook.com/terms.php.

[6] We do not claim that Facebook did not try to make its policy more community-oriented. We argue that it did not select a way that would involve the community

To achieve prosocial governance and a vital online community the relationship between users and platforms must shift from one focused on platform authority and user obedience to an environment focused on user internalization of rules regarding content and its moderation. In the case of voluntary rule-following, the traditional model encourages people to hide their behavior and requires authorities to search for rule-breaking. This is a challenging task. When people are invested in following the rules, they view doing so as a personal obligation and do it without reference to whether their conduct can be observed and sanctioned.[7] While online platforms may seem adept at monitoring their users' behavior, in reality, they have found that people find creative ways to hide.[8] For example, they may create multiple accounts and fake identities. Social platforms have inevitably been thrown into a role familiar to police departments: watching their community and trying to identify violations.

---

of users effectively and create a more bottom-up approach. Also despite its efforts over the years to publish its policies, the implementation of those policies remained obscure for the day-to-day user. For the changes in Facebook content governance approach over the years, see Rotem Medzi, *Enhanced self-regulation: The case of Facebook's content governance*, NEW MEDIA & SOC'Y (2021). In 2009, Facebook announced that it planned to try new forms of governance and giving more authority to the users. The problem with that approach, which later failed, was that it did not allow the users to self-regulate, and it was not very clear how Facebook enforced those policies. It only allowed the users to vote on the changes that Facebook had undertaken. *Facebook Opens Governance of Service and Policy Process to Users,* FACEBOOK (Feb. 26, 2009), https://about.fb.com/news/2009/02/facebook-opens-governance-of-service-and-policy-process-to-users/. This new governance approach failed since not many participated in voting. *Results of the Inaugural Facebook Site Governance Vote***,** FACEBOOK (Apr. 24, 2009), https://web.archive.org/web/20090430215524/http://blog.facebook.com/blog.php?post=79146552130. In 2012, Facebook decided to remove the voting mechanism altogether and instead reach out to a select number of third-party experts. Elliot Schrage, *Proposed Updates to our Governing Documents***,** FACEBOOK (Nov. 21, 2012), https://about.fb.com/news/2012/11/proposed-updates-to-our-governing-documents/.

[7] *See* TOM R. TYLER, WHY PEOPLE OBEY THE LAW (2006).

[8] Lauren Reichart Smith et al., *Follow Me, What's the Harm: Considerations of Catfishing and Utilizing Fake Online Personas on Social Media*, 27 J. LEGAL ASPECTS SPORT 32 (2017).

This approach to motivating rule compliance should be familiar to anyone who works in the criminal justice system or understands how it works. It is an approach based upon the idea that people will follow rules or laws because they fear the consequences of failing to do so and that in order to ensure that people do follow rules, the punishment or the threat of punishment must be severe enough to motivate a rational actor to follow the rules. Two of us, Professor Meares and Professor Tyler, have spent the past two decades explaining the ways in which this approach to compliance in criminal law does not work well and often in fact undermines the stated goals of the system.[9] We have argued in favor of approaches that encourage internalization of rules based on enhancing citizen trust in legitimacy of various kinds of authorities.[10] We characterize these approaches as prosocial in that their goal is to promote and enhance existing positive norms of behavior as opposed to making central the ferreting out and punishing of bad behavior. In this paper, we apply ideas we have developed in the criminal justice space to online platforms,[11] and we theorize that prosocial governance

---

[9] TYLER, *supra* note 7; TOM R. TYLER, WHY PEOPLE COOPERATE: THE ROLE OF SOCIAL MOTIVATIONS (2013) [hereinafter WHY COOPERATE]; Tom R. Tyler, *Enhancing Police Legitimacy*, 593 ANNALS AM. ACAD. POL. & SOC. SCI. 84 (2004) [hereinafter *Police Legitimacy*]; Tom R. Tyler & Tracey L. Meares, *Procedural Justice Policing*, *in* POLICE INNOVATION: CONTRASTING PERSPECTIVES 71 (David Weisburd & Anthony Braga eds., 2019); Tracey L. Meares, *The Path Forward: Improving the Dynamics of Community–Police Relationships to Achieve Effective Law Enforcement Policies*, 117 COLUM. L. REV. 1355 (2017); MEGAN QUATTLEBAUM ET AL., JUSTICE COLLABORATORY AT YALE L. SCH., PRINCIPLES OF PROCEDURALLY JUST POLICING (2018); Tracey L. Meares et al., *Lawful or Fair? How Cops and Laypeople Perceive Good Policing*, 105 J. CRIM. L. & CRIMINOLOGY 297 (2015); TOM R. TYLER, LEGITIMACY AND CRIMINAL JUSTICE: AN INTERNATIONAL PERSPECTIVE (2007); Tom R. Tyler, *What Is Procedural Justice—Criteria Used by Citizens to Assess the Fairness of Legal Procedures*, 22 LAW & SOC'Y REV. (1988) [hereinafter *Procedural Justice*].

[10] Tracey L. Meares & Tom R. Tyler, *Justice Sotomayor and the Jurisprudence of Procedural Justice*, 123 YALE L.J. F. 525 (2014).

[11] In what follows, we frequently refer to "platforms." Platforms are a means by which people can engage with one another through a network, including the

approaches can contribute to community vitality online. We think that prosocial governance approaches encourage people to follow platform rules and internalize rule-following, and more importantly, to engage with problems and cooperate to solve them constructively. Both of these goals are important, though they may have different ends. The first is good for platforms and their operation. The second is good for society. Since these are complementary ends, we treat them as equal goals.

Platforms must also engender cooperative engagement among community members who use the platform as a forum to address common problems constructively. Community members' motivations for creating a cooperative space are several. First, platforms want their users to enjoy their time on the platform and find it both a positive experience and one that is useful to them in managing the problems in their lives. This cooperative engagement is also important because it enhances the capacity of communities to work together and thereby improves social, economic, and political well-being. When people communicate in positive and constructive ways, they are better able to work together to address common issues and problems.[12] When people are better able to work together, they create stronger communities because they can and do address the needs in those communities more effectively.

Internet. Platforms are always controlled by a single entity. In particular, the function of the platform is subject entirely to the control of the platform operator, so to get access to the functionality offered by the platform, the users must accept the platform's terms of service. Platforms are usually accessed through the World Wide Web but need not be. This use of "platforms" does not include software development platforms or other such uses common in the tech industry; it is primarily societally defined and comprises at least everything popularly described to be a "social media platform," but also includes systems that are often not thought of as social media (such as GitHub or Stack Overflow).

[12] Tom R. Tyler & Steven L. Blader, *The Group Engagement Model: Procedural Justice, Social Identity, and Cooperative Behavior*, 7 PERSONALITY & SOC. PSYCHOL. REV. 353 (2003).

Just as the criminal justice system is a way of governing human interaction in the offline world, there are ways of governing human interaction in the online world.[13] Some of those governance methods depend on similar deterrence strategies to those that have been used in the criminal justice system,[14] so there is reason to believe that reform strategies that apply to the criminal justice system will apply to communities and governance methods online. It is commonplace to use authority-based governance (which depends upon sanctions and deterrence) as opposed to community-based governance (which depends upon willing consent) to manage online behavior.[15] Authority-based governance operates through the rules, practices, and procedures adopted by social media platforms and their employees—decision-makers such as content reviewers, policymakers, and product designers. Authority-based governance is built from the norms and values of the platform and not those of the community it serves.[16] By contrast, in community governance,

---

[13] A very detailed account of how social media platforms and social networks govern their users can be found in Danah M. Boyd & Nicole B. Ellison, *Social Network Sites: Definition, History, and Scholarship*, 13 J. COMPUT.-MEDIATED COMM. (2007); Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598 (2017).

[14] Scholars have mentioned the similarities between platform governance and punitive governance in more detail. We have mentioned these sources in *infra* note 42. To make the similarities more tangible we provide a few punitive approaches here: Tinder (a dating app) permanently bans its users if they violate the rules and is quite unforgiving. Tinder's policy says, "If you violate any of these policies, you might be banned from Tinder. Seriously, don't make us Swipe Left on you—because ***there will be no do-overs once we do.***" *Community Guidelines*, TINDER, https://policies.tinder.com/community-guidelines/intl/en/ (emphasis added) (last visited Mar. 20, 2021). Twitter's enforcement actions are punitive too and can escalate: Twitter can limit tweet visibility and require tweet removal, hide a violating Tweet while awaiting its removal, place an account in read-only mode, and permanently suspend an account. *Our Range of Enforcement Options*, TWITTER, https://help.twitter.com/en/rules-and-policies/enforcement-options (last visited Mar. 20, 2021).

[15] Thomas C. O'Brien et al., *Building Popular Legitimacy with Reconciliatory Gestures and Participation: A Community-Level Model of Authority*, 14 REGUL. & GOVERNANCE 821 (2020).

[16] In Bradford et al.'s transparency report, authority-based governance is presented as top-down governance. In this governance model, the social media

the community—a group of people that share common goals and interests—helps to make and enforce the norms, procedures, and practices by which the platform is governed.

Our goal in this paper is to advance a theory of a self-motivated prosocial production system—that is, a system that by its nature produces a cycle of socially desirable inputs. Research demonstrates that process-based fairness rooted in social psychology is a promising approach.[17] Procedural justice requires affording the community a voice and opportunities for participation, the use of neutral procedures for decision-making, treatment with respect and dignity, and communication of trustworthy motives through consideration of and responsiveness to people's needs and concerns. Decision-makers and community members can generate prosocial behavior over the long term by adhering to the principles of procedural justice. In effect, there are two goals for our prosocial production system: to limit negative experiences and to promote positive behavior.

A key factor in achieving prosocial engagement online, though, is alignment with platform business models. The business models are heavily metric- and product-driven.[18] One reason why platforms focus on identifying bad behavior and then demonstrating a particular consequential response is that those actions are easy to

---

platform issues a detailed set of rules that leaves little opportunity for the community of users to come up with their own rules. BEN BRADFORD ET AL., JUSTICE COLLABORATORY AT YALE L. SCH., REPORT OF THE FACEBOOK DATA TRANSPARENCY ADVISORY GROUP 31 (2019).

[17] Tyler and Meares have illustrated this in Tracey L. Meares & Tom R. Tyler, *Justice Sotomayor and the Jurisprudence of Procedural Justice*, 123 YALE L.J.F. 525 (2014).

[18] As stated by Venkatesh, many tech companies have adopted the myth of "product is governance." Working based on this myth, companies have deemed reliance on self-regulation by the users as inefficient for governance because of the volume and scale of content that needs to be governed. Sudhir Venkatesh, *The Myth of Platform Governance: How Product Culture Shapes Content Moderation in Technology Firms*, YALE J. L. & TECH. (Forthcoming 2021).

measure—not unlike arrests in the real world.[19] Those who govern the behavior of online communities (mainly platforms' employees and policymakers who are engaged with user and content moderation) will want to determine to what extent procedural justice has in fact helped achieve prosocial goals. Thus, another contribution of this paper is to provide potential benchmarks for measuring the success we think our theoretical approach can achieve. We offer an explanation for how platforms could create an experiment to measure any prosocial approach to community vitality.[20]

In summary, then, the overall approach can be depicted in the following table:

GOALS	**ACTIVITIES**	
	*On the **platform***	*In the **community***
Limit negative experience →	Content moderation to lessen the amount of hate speech, false content, etc. →	Lessen the impact of negative and false content on real world activities such as elections
Promote positive behavior →	Encourage constructive interactions →	Increase the resilience and vibrancy of real-world communities

[19] The report of the Facebook Data Transparency Working Group shows this similarity by drawing an analogy between "Facebook's prevalence measurement" and "commonly used measures of crime." BRADFORD ET AL., *supra* note 16, at 18-19.

[20] Platforms do commonly survey their users to test for what is often called "customer satisfaction" with the platform experience (user experience or UX) or feelings about their experience. These studies are typically not considered within a framework of the site's capacity to create positive social experiences with an eye to community building. The performance metric matters because time on the site is a profit indicator. In a business sense, time on a platform spewing hate speech and time promoting tolerance are both related to platform business model success.

Our argument is expanded below. Section II discusses how platform governance has evolved and led to the adoption of deterrence-based approaches. Section III theorizes a prosocial system that platforms and online communities could use to limit antisocial behavior, promote prosocial behavior, and potentially measure the effect of the prosocial system on their platforms. Section IV concludes the paper.

GOVERNING ONLINE BEHAVIOR

Our argument depends on an analogy to how the criminal legal system effectively and fairly addresses criminal behavior in the real world. In section A, we discuss the gradual change from community- to authority-based governance on platforms. Section B lays out the shortcomings of deterrence-based governance approaches on platforms, which are often authority-based, by drawing an analogy between such methods and criminal justice methods.

Emergence of Online Norms and Governance of Platforms

It is easy to imagine that developing and conforming to online terms of service is a straightforward matter, but in fact, online behaviors and the norms governing them developed gradually and under the influence of multiple social and legal pressures. It is useful to divide the periods of online platform governance into different phases. The first is one of community governance, which came about in the early 1990s, before the Internet became the overwhelmingly dominant communication mechanism. During this time, online communities with a wide range of interests and goals sprung up, but they relied on different technologies with their own

social affordances,[21] and volunteers and system administrators ran various virtual spaces.[22] These online communities used the Internet and other communications technologies to achieve their goals,[23] which ranged from discussing their favorite shows, politics, and literature to organizing political gatherings, holding community meetings, and solving each neighborhoods' problems.

The Internet as a whole is broadly decentralized, and these early Internet-based communities usually governed themselves in a decentralized manner. Prior to the emergence of the web, there were various ways in which people gathered online in communities. Two prominent ones were Listservs (sometimes called "mailing lists" or just "lists," terms still sometimes in use today) and Usenet newsgroups (Usenet started outside the Internet and was in wide use for several years but has mostly ceased to play a role in people's

---

[21] Barry Wellman et al., *The Social Affordances of the Internet for Networked Individualism*, 8 J. COMPUT.-MEDIATED COMM. (2003); Laura W. Black et al., *Self-Governance through Group Discussion in Wikipedia: Measuring Deliberation in Online Groups*, 42 SMALL GROUP RES. 595 (2011).

[22] These distributed systems were run by system administrators, who managed their technical maintenance needs. As these communities grew, their system administrators did not want to get involved with governance and so asked the communities themselves to take part in decision-making. For example, platforms like LambdaMOO and Habitat made major changes to their governance and used community governance mechanisms such as "grassroots petitions" and "collective voting." Sherry Turkle, *Virtuality and Its Discontents: Searching for Community in Cyberspace*, *in* THE WIRED HOMESTEAD: AN MIT PRESS SOURCEBOOK ON THE INTERNET AND THE FAMILY 385 (Joseph Turow & Andrea L. Kavanaugh eds., 1996). LambdaMOO and Habitat were early online multi-user environments where people interacted with each other through pre-web technologies. Diane J. Schiano, *Lessons from LambdaMOO: A Social, Text-Based Virtual Environment*, 8 PRESENCE: TELEOPERATORS & VIRTUAL ENV'T 127 (1999); Chip Morningstar & F. Randall Farmer, *The Lessons of Lucasfilm's Habitat*, 1 J. VIRTUAL WORLDS RES. (2008).

[23] Black et al., *supra* note 21; HOWARD RHEINGOLD, TOOLS FOR THOUGHT: THE HISTORY AND FUTURE OF MIND-EXPANDING TECHNOLOGY (2000); MARC A. SMITH & PETER KOLLOCK, COMMUNITIES IN CYBERSPACE § 1 (1999); Constance Elise Porter, *A Typology of Virtual Communities: A Multi-Disciplinary Foundation for Future Research*, 10 J. COMPUT.-MEDIATED COMM. (2004).

online experience).[24] The nature of the technical operation of those technologies meant that multiple Internet site operators had some role to play.[25] It was not possible to take control of the operation of a site (a newsgroup, for example). These forums, mostly based on Listservs and Usenet newsgroups but sometimes on the early web, were distributed and decentralized in operation (even if, as in some cases, they depended on centralizing technology like the web). The systems were for the most part technically basic, so they depended on cooperative administration. Attempts to impose central control resulted in people objecting by setting up alternatives.[26]

The second phase started in the late 1990s, when the web became very popular. The Internet and the World Wide Web (often just called "the web") are not the same technology, and the difference may influence the governance models that emerge in each system. The Internet is a global network made up of many independent, globally interconnected networks. As online platforms grew, the networks specialized more or started using the more centralized technology of the web. However, early adopters of the web (such as Wikipedia and Slashdot) used "community

---

[24] Usenet was a global bulletin board that allowed user-to-user interaction through their local news servers. Users would send messages from their server to other users' servers, and they could communicate and react to each message. It is important to note that the web was technically distributed, but it also allowed for centralized governance. For example, it could turn the website operator into an exclusive intermediary (a service provider) because the operator could disallow user-to-user interaction, and the users had to communicate through the website. An example can clarify this: if Facebook removes a group from its website, the members no longer have access to that group under any circumstances. But if a server no longer hosts a newsgroup, the users could move to another server and have access to the same newsgroup. Bryan Pfaffenberger*, A Standing Wave in the Web of Our Communications*: *Usenet and the Socio-Technical Construction of Cyberspace Values*, *in* FROM USENET TO COWEBS 20 (Christopher Lueg & Danyel Fisher eds., 2003).

[25] PETER H. SALUS, CASTING THE NET: FROM ARPANET TO INTERNET AND BEYOND (1995).

[26] *Id*. at 144.

governance" mechanisms akin to those of older systems.[27] Only gradually did social media platforms adopt a hybrid authority-community governance or a more hierarchical, authority-based governance.[28] Before the emergence of centralized platforms, the typical virtual space was like a main street where communities grew. As platforms used technology that tended to encourage centralized operation and control, virtual spaces became more like shopping malls, and users turned into customers.[29]

The third phase started in the mid-2000s. Scholars warned that when the commercial stakes in online communities rose, so too would the interest in directing the participants' attention or controlling the format of interaction to suit the profit-making agendas of corporate partners.[30] It was around 2006 that the commercial stakes became high once certain platforms began amassing users and generating revenue by using their online platforms to regulate user behavior, rather than just facilitating communication. Some became multisided online markets, providing services other than facilitating communication. Economically, it was in these platforms' interest to keep users inside their "ecosystems." Examples of this pattern include some of the most familiar names in online platforms, such as Facebook, Twitter, or

---

[27] See the following articles for a more detailed account of Wikipedia and Slashdot governance mechanism: Cliff Lampe & Paul Resnick, *Slash(dot) and Burn: Distributed Moderation in a Large Online Conversation Space*, *in* PROCEEDINGS OF THE SIGCHI CONFERECE ON HUMAN FACTORS IN COMPUTING SYSTEMS 543 (2004); Aleksi Aaltonen & Giovan Francesco Lanzara, *Building Governance Capability in Online Social Production: Insights from Wikipedia*, 36 ORG. STUD. 1649 (2015); Laura Stein, *Policy and Participation on Social Media: The Cases of YouTube, Facebook, and Wikipedia*, 6 COMM. CULTURE & CRITIQUE 1 (2013).
[28] BRADFORD ET AL., *supra* note 16.
[29] Turkle, *supra* note 22.
[30] BRADFORD ET AL., *supra* note 16, at 30–31; Dara Byrne, *The Future of (the) 'Race': Identity, Discourse, and the Rise of Computer-mediated Public Spheres*, *in* LEARNING RACE AND ETHNICITY, YOUTH AND DIGITAL MEDIA (Anna Everett ed., 2008). Byrne explained that "As the commercial stakes in online communities rise, so too will the interest in directing the attention of participants, or controlling the format of interaction, to suit the profitmaking agendas of corporate partners."

Sina Weibo, but an exhaustive list is now impractical because the pattern is so widespread. The incentive to keep users in their "ecosystem" meant that, unlike pre-web systems, these newer platforms developed technical features that were only available inside that platform, so alternatives were not possible.

In platforms with authority-based governance, users went from being members of a particular online community to being subjects of the platform. This change in governance structure might have been because, as platforms' networks became larger, they did not think it feasible to leave their communities to govern themselves.[31] But it also might have been because the platforms' interests were better served by their power over their users.

This is not to say that "community governance" does not exist on online platforms anymore. Online platforms might adopt hybrid governance mechanisms that use several mechanisms for governing the behavior of users:

1. A top-down user agreement and a content moderation policy drafted by the platform's lawyers;
2. Community rules that the community generates within its various sub-groups;
3. Overall community rules (Netiquette, Reddiquette, or the like) which are not binding, but to which the community as a whole contributes and offers amendments.

Some might argue that Facebook and other centralized platforms are investing in features and policies that can empower their communities.[32] This is true, yet they still have dominant

---

[31] BRADFORD ET AL., *supra* note 16, at 30.

[32] For example, Facebook creates a sense of community by "group building." It states that "Facebook gives you powerful tools to help your group thrive. These

authority-based governance in place. For example, Facebook empowers its community to convene various groups and set up their own rules and code of conduct. But the community (members of the group) does not have much say in policy changes. Facebook has well-elaborated community standards (mainly drafted by Facebook lawyers) that impose restrictions on many aspects of individuals' and communities' behavior. It does not leave much room for self-governance.[33]

A better example of the hybrid model is Reddit, which has its own terms and conditions and imposes standards of behavior but also allows communities of users to assert their own rules. Reddit emphasizes the community and the role of the moderators, explaining that it rarely wants to get involved with content moderation: "Reddit may, at its discretion, intervene to take control of a community when it believes it in the best interest of the community or the website. This should happen rarely (e.g., a top moderator abandons a thriving community), but when it does, our goal is to keep the platform alive and vibrant, as well as to ensure your community can reach people interested in that community."[34] Hence Reddit does intervene, but the basic interaction is first within

---

focused tutorials give you more information on these helpful features and how to use them." *Using Key Group Tools,* FACEBOOK, https://www.facebook.com/community/using-key-groups-tools/ (last visited Jan. 24, 2021).

[33] BRADFORD ET AL., *supra* note 16, at 31. In a study about Facebook, YouTube, and Wikipedia, Stein explained that users did not know about Facebook policy changes until they came into effect, and they challenged Facebook's policies about issues such as privacy. She concluded that (at the time of writing the paper) Facebook and YouTube gave their users minimal control over content and governance of the website. Stein, *supra* note 27, at 354. It is of note that Facebook undertakes meetings with third-party experts to discuss its policies and might make changes accordingly, but third-party experts might not be community members. One of the first consultations of this kind happened in 2012 at Stanford. See Klonick, *supra* note 13.

[34] *Moderator Guidelines for Healthy Communities,* REDDIT, https://www.redditinc.com/policies/moderator-guidelines (last visited Jan. 20, 2021).

the community, and the platform rules are secondary. Reddit also has an "informal expression of value"[35] called Reddiquette that communities refer to, in addition to the formal terms and conditions provided by Reddit.[36] Because Reddiquette is a normative system based on Reddit users' values, it is more acceptable to users than top-down platform rules.[37] So, Reddit offers examples of all three modes of the hybrid governance mechanism at once.

Another example of a hybrid governance mechanism is Wikipedia. Wikipedia's policies mainly come from its community of editors. Similar to Reddit's Reddiquette, English-language Wikipedia has a Wikiquette. Community editors have written down behavioral standards that can be changed by community members through consensus.[38] However, Wikipedia also has a top-down mechanism that involves its legal team. That mechanism can be invoked to make decisions that overrule the community. For example, their trust and safety group can govern its users' behavior. The group's Wiki page indicates that it "aims to defer to local and global community processes to govern on-wiki interactions." While acknowledging that intervention may happen rarely, it also states that they step in to protect the safety and integrity of users, contributors, and the public.[39]

---

[35] *Reddiquette,* REDDIT (2020), https://reddit.zendesk.com/hc/en-us/articles/205926439-Reddiquette.

[36] *Content Policy*, REDDIT, https://www.redditinc.com/policies/content-policy (last visited Jan. 20, 2021).

[37] Fiesler et al. illustrated this by undertaking empirical research and concluded, "It is also much more common for subreddits to refer to Reddiquette than official policy, suggesting again that the rules closest to the community itself are the most visible, prioritizing an individual subreddit over Reddiquette over Reddit policy." Casey Fiesler et al., *Reddit Rules! Characterizing an Ecosystem of Governance*, *in* 12TH INTERNATIONAL AAAI CONFERENCE ON WEB AND SOCIAL MEDIA 78 (2018).

[38] *Etiquette,* WIKIPEDIA, https://en.wikipedia.org/wiki/Wikipedia:Etiquette (last visited Jan. 20, 2021).

[39] *Trust and Safety,* WIKIPEDIA, https://meta.wikimedia.org/wiki/Trust_and_Safety (last visited Jan. 20, 2021).

Repeating the Same Mistake Online

The criminal legal system relies heavily on sanctions or threat of them to achieve behavioral governance goals. The three-strikes laws and mandatory minimum sentences and guidelines to increase penalties for certain crimes are but two examples of applying mandatory sanctions to certain behaviors.[40] As we have explained, this approach to governing people's behavior centers deterrence as a theory of compliance. Deterrence-based approaches focus on increasing the cost of rule-breaking so that people, out of self-interest and fear of punishment, do not break the rules. These approaches have notable weaknesses, however. They are most effective in situations where surveillance is possible, and because they depend so much on surveillance, in the real world they can be extremely costly.[41]

In the realm of social media, governance mechanisms and the popular rules that platforms usually implement also rely on deterrence. These systems focus mainly on compliance and individual violations. Their common punitive measures are analogous to those in the criminal justice system. Platforms typically suspend accounts in the face of infractions—the functional equivalent to putting someone in "jail," which can operate to incapacitate a person or punish them or both. Sometimes, after multiple violations, a platform might ban a user from the platform entirely (analogous to so-called "three strikes" laws).[42]

---

[40] Andrew V. Papachristos et al., *Why Do Criminals Obey the Law? The Influence of Legitimacy and Social Networks on Active Gun Offenders*, J. CRIM. L. & CRIMINOLOGY 401 (2012).

[41] TYLER, *supra* note 7, at 263; Tracey Meares, *Broken Windows, Neighborhoods, and the Legitimacy of Law Enforcement or Why I Fell in and out of Love with Zimbardo*, 52 J. RES. CRIME & DELINQ. 609 (2015).

[42] Scholars have drawn an analogy between mechanisms of the criminal justice systems and these platforms' governance approaches in the past. This might especially be because the line between digital and real life has been blurred, and

Content moderation as it operates today is similar to focusing on "arrest rates" and "crime rates" in the criminal legal system.[43] Both mechanisms are outcome-oriented and do not have as goals either prevention or reform. Rather than attempting to change users' behavior through education or even mere notice of the rules, the major focus of many platform moderation efforts is simply to count and reduce individual violations. These "elimination" measures do not positively contribute to users' behavior; for example, they do not encourage users not to repeat the offense. However, Tyler et al. undertook an experimental study about Facebook which empirically showed that the users that Facebook treated fairly during content moderation were more likely not to repeat the offense than those who were not treated fairly.[44]

There is a link between the growth and change of structures for online interaction and the mode of governance. As the Internet grew and main streets turned into shopping malls, platforms increased their use of aggressive methods like content takedowns and blocking and suspending accounts. We contend that in shifting from communal governance approaches to more authority-based ones, platforms started making the same mistakes that the judicial system and the police now make: focusing on individuals and

---

users' lack of access to these platforms might highly affect their access to their community, families, and friends. It can even give them the feeling that their access to their community was cut off as a result of a platform's suspension or ban. Tyler et al. argued that account suspension or cancellation "parallels" some criminal justice mechanisms, such as incarceration. Tyler et al., *supra* note 2. Other scholars have also mentioned that platforms' governance approaches are punitive, and they tend to adopt or are more likely to use methods similar to criminal justice. Platforms' inclination toward punitive governance approaches is elaborated in the following sources: Sarah Myers West, *Censored, Suspended, Shadowbanned: User Interpretations of Content Moderation on Social Media Platforms*, 20 New Media & Soc'y 4376 (2018); Sarita Schoenebeck et al., *Youth Trust in Social Media Companies and Expectations of Justice: Accountability and Repair After Online Harassment*, 5 PACM ON HUM.-COMPUT. INTERACTION (2021).

[43] BRADFORD ET AL., *supra* note 16, at 19.

[44] Tyler et al., *supra* note 2.

eliminating or weakening the communities with authority-based governance. This is of course not to say that these efforts have so far led to the pernicious effects of what we see in the criminal justice system. Importantly, the web itself is only thirty years old, and online experience is still but a fraction of human life. Our point is simply this: It is unwise to continue to build models of online governance founded on assumptions similar to those that have had poor effects in criminal justice. Since better approaches for criminal justice have already been proposed, perhaps those models can also be applied effectively in the online world.

IMAGINING PROSOCIAL PRODUCTION ON PLATFORMS

So what does a good alternative governance approach on platforms look like? In this part, we argue that focusing on community vitality as a goal rather than merely identifying and punishing bad behavior motivates the production of prosocial governance mechanisms that can facilitate compliance, engagement, and cooperation. Relying on the theory of procedural justice, we conceptualize a prosocial production system that could lessen violations on platforms by motivating individuals to internalize rules, voluntarily comply with them, and engage in healthful interaction.

Before focusing on community vitality and applying procedural justice, we cover several important work that scholars and practitioners have done in this space. Eli Pariser, a tech-entrepreneur, has undertaken several initiatives that focus on communities on social media platforms. For example, Civic Signal (new public) works on creating "public spaces" on platforms.[45] They also work on creating vibrant, livable online spaces. One of the

[45] NEW PUBLIC, https://newpublic.org/ (last visited June 14, 2021).

inspiration for this work is Jane Jacobs, an urbanist and activist who objected to the elimination of communities and social structure. She also fought against building highways and fake parks in the suburbs at the expense of demolishing communities.[46] Drawing an analogy between Jacob's work on neighborhood and communities, digital activists and engineers have tried to envision social structures on platforms.[47]

As we mentioned, procedural justice in decision-making is central to creating vital communities. Research demonstrates that four factors matter.[48] The first is participation or voice: the decision-maker should give people the opportunity to explain their situation and perspective.[49] Participation in decision-making processes should happen at various stages. This means that people's voices should not exclusively be heard after a dispute arises, but they should be able to take part in different stages of decision-making processes that can affect them. This can be during any or all of policy-making, dispute resolution, or enforcement processes. Second, people care about being able to ascertain whether authorities are being fair as they carry out decisions. Fairness includes the following: neutrality, objectivity, factuality of decision-making, consistency in decision-making, and transparency.[50] Third, people care a great deal about being treated with dignity and respect. People care about how their community leaders and authorities treat

---

[46] JANE JACOBS, THE DEATH AND LIFE OF GREAT AMERICAN CITIES (2016).
[47] Amy X. Zhang et al., *PolicyKit: Building Governance in Online Communities*, *in* PROCEEDINGS OF THE 33RD ANNUAL ACM SYMPOSIUM ON USER INTERFACE SOFTWARE AND TECHNOLOGY 365 (2020). Zhang et al have come up with a software design that focuses on designs that could potentially lead to vital communities and plurality in governance.
[48] *Procedural Justice*, *supra* note 9, at 103; BRADFORD ET AL., *supra* note 16.
[49] POLICE LEGITIMACY, supra note 9.
[50] Tom R. Tyler & Cheryl J. Wakslak, *Profiling and Police Legitimacy: Procedural Justice, Attributions of Motive, and Acceptance of Police Authority*, 42 CRIMINOLOGY 253 (2004).

them; they usually respond positively to being treated with dignity, respect for their rights, and politeness.[51] Finally, people want their leaders and decision-makers to act out of a sense of benevolence toward them, so it is important that they perceive authorities to be communicating trustworthy motives. People attempt to discern why authorities act the way they do, and a procedurally just decision-making process gives them the perception that the authorities are benevolent, well-intentioned, and sincere, and do not act only out of self-interest.[52]

Procedural justice is central to the creation of a self-motivated prosocial production system—that is, a system that by its nature produces a cycle of socially desirable inputs. In the following sections, we argue that platforms can motivate voluntary rule-following through procedural justice. Importantly, members of online communities can cooperate with each other and the authorities to lessen the impact of rule violations and negative behavior. We also argue that this approach can do more than motivate rule-following. We think prosocial approaches are an additional step platforms can take to encourage their communities to actively do good.

Prosocial Compliance and Cooperation: Limiting Antisocial Behavior on Platforms

The traditional goal of content regulation is to avoid harm by limiting negative content that violates platform rules. Prosocial approaches begin with the goal of limiting negative content but are also concerned with the objective of promoting positive content. Prosocial approaches treat a positive social environment as a goal.

---

[51] *Id*. at 253.
[52] Tom R. Tyler & E. Allan Lind, *A Relational Model of Authority in Groups*, 25 ADVANCES EXPERIMENTAL SOC. PSYCHOL. 115 (1992).

The goal of achieving a positive environment has two aspects. The first is to aid traditional regulation. When platforms enforce their rules through traditional control mechanisms, they identify and sanction undesirable content. This motivates users to evade platform authorities and hide their actions. However, when users identify with and feel positively about the provider and their online community, they become more self-regulatory. To put it simply, they are more likely to want to do the right thing and to do it voluntarily. Hence, building a positive online climate facilitates effective regulation. An important part of this positive climate is accepting the legitimacy of the platform, its rules, and its enforcement mechanisms. When legitimacy is high, the threat of sanctions in not the primary means of promoting rule adherence.

The second goal is for the platform to serve as a safe space within which the members of different communities can interact constructively and civilly to address their common issues and concerns. This positive climate will make the time that people spend on a platform more satisfying and will also enhance the possibility of useful dialogue about potentially divisive issues. That dialogue can then spill over into real-world communities and enable them to jointly address their political, social, and economic issues. Again, the legitimacy of the platform as an "honest broker" seeking to create a secure and safe space for such discussions is crucial. A key antecedent of legitimacy is the belief that an authority is benevolent and sincere, seeking in good faith to help people define and address their needs and concerns.[53]

---

[53] See Tom Tyler, *Policing in Black and White: Ethnic Group Differences in Trust and Confidence in the Police*, 8 POLICE Q. (2005).

Tyler defines legitimacy as the belief that authorities have the right to dictate proper behavior.[54] Meares defines legitimacy as a collection of individuals' perceptions of the laws and the authorities that enforce them.[55] People comply with the law as long as they perceive the authorities and their laws as legitimate. When people perceive authorities as legitimate, they will largely regulate their own behavior, so hierarchical enforcement mechanisms will be less necessary.[56] By resorting to procedural justice (instead of deterrence-based mechanisms) and building the legitimacy of decision-makers, it is possible to encourage people to comply with the rules.

To limit antisocial behavior on platforms, we need to go beyond compliance and address cooperation. Cooperation includes the willingness to accept authority, deference to the decisions made by the authority, and everyday rule adherence. Cooperation is also the willingness to aid decision-makers (the authorities in a governance mechanism) in identifying violations and wrongdoers and helping with the adjudication of conflicts. Tyler and Jackson demonstrated that cooperation can be achieved by trust and confidence in an authority; it can also be achieved by normative alignment, or sharing the authorities' goals and values.[57]

A crucial question is, how can platforms generate compliance and cooperation? Recall that procedural justice suggests that people are more likely to follow rules when they participate in

---

[54] Tom R. Tyler, *Procedural Justice, Legitimacy, and the Effective Rule of Law*, 30 CRIME & JUST. 283 (2003).
[55] Tracey L. Meares, *Norms, Legitimacy and Law Enforcement*, 79 OR. L. REV. 399 (2000).
[56] Tom Tyler & Steven Blader, *Can Businesses Effectively Regulate Employee Conduct? The Antecedents of Rule Following in Work Settings*, 48 ACAD. MGMT. J. 1143 (2005).
[57] Tom R. Tyler & Jonathan Jackson, *Popular Legitimacy and the Exercise of Legal Authority: Motivating Compliance, Cooperation, and Engagement*, 20 PSYCH. PUB. POL'Y & L. 78 (2014).

the decision-making process and feel that the decision-maker has heard their voice. Translating this participation into the online environment can take various forms. For example, community groups might come up with their own rules, or community members might get meaningful participation in the policy-making process, or community members might receive fair treatment during the adjudication process. Establishing a platform's legitimacy might be harder than it is for other authorities that have been approved by their community members and are appointed through a democratic process. Platforms' decision-makers that adjudicate disputes and enforce rules are the employees of the platform and not selected or appointed through a democratic process. This might affect users' incentives to follow rules, since they might not buy into the outcome of the adjudication, so they might try to subvert the norms and the outcomes that the platform tries to enforce.

When authorities over a social group treat group members fairly, the members feel included, find the group valuable and valid, and identify with its values.[58] The fair treatment preference is constant in various settings and different communities. Diversity in ethnicity, location, and other aspects does not usually affect individuals' preference for fair treatment.[59] This is especially important in the context of platforms because they serve global and diverse communities. This is not to say that fairness of their

---

[58] WHY COOPERATE, *supra* note 9.

[59] Tyler and Huo examined whether race or ethnicity has an impact on authorities' personal perception that can weaken procedural justice generality. TOM R. TYLER & YUEN J. HUO, TRUST IN THE LAW 153 (2002). Other scholars have tested the generality of procedural justice and its applicability to various settings. So far, a couple of studies have come to the conclusion that procedural justice leads to cooperation and compliance across different settings. Scott E. Wolfe et al., *Is the Effect of Procedural Justice on Police Legitimacy Invariant? Testing the Generality of Procedural Justice and Competing Antecedents of Legitimacy*, 32 J. QUANTITATIVE CRIMINOLOGY 278 (2016); JONATHAN JACKSON ET AL., JUST AUTHORITY? TRUST IN THE POLICE IN ENGLAND AND WALES (2012).

treatment has an absolute effect on people. There are circumstances under which fairness might not motivate cooperation.

Fair treatment requires decision-makers to be objective and neutral. On social media platforms, we can detect fairness or unfairness during the enforcement process. However, to go beyond applying procedural justice to dispute resolution and enforcement processes, it is important to consider the fairness of interactions between platform decision-makers and platform users as well as among community members. Fairness of interactions (for example, showing tolerance to opposing views and considering all arguments based on merit) can lead to building a community with members that perceive the processes and decision-makers to be fair, which leads to further cooperation within the community.

Offline or online, people care about being treated with dignity and respect during interactions, whether with other community members or with the decision-makers on the platform.[60] Elimination of disrespectful content does not in itself afford people such respect. However, an increase in respectful treatment can provide people with what they desire and also provide a chance to cooperate with authorities.

Communicating trustworthy motives might be especially important in the case of unelected authorities, whether they are platform owners or community leaders who are not elected by the community members. Commercially driven initiatives and their commercially driven authorities, especially, should make sure not to communicate only profit-making incentives when they make and enforce decisions that affect the community. To be effective, they should have the best interests of the community in mind, avoid

---

[60] In our paper about Facebook, we demonstrated that users in an online setting care about procedural fairness. Tom Tyler et al., *supra* note 2.

acting merely out of self-interest, and communicate all of that effectively.

We can see elements of participation especially on some platforms with hybrid governance models, since they allow their users to participate in decision-making processes. For example, Nextdoor (a neighborhood social media platform) allows the community member volunteers, neighborhood leads, and group admins to make decisions and enforce Nextdoor's guidelines.[61] The users also get to vote about whether to remove a given piece of content. When the votes pass a certain threshold, the lead for the neighborhood takes the content down.[62] This is a good way to get people to cooperate with the authorities of groups—in Nextdoor's case, the leads of the neighborhood.

To some extent, it is possible to compensate for the shortcomings of top-down rules by being procedurally just. As Tyler et al. showed in their paper on Facebook's governance model, when rule violators on Facebook were treated with procedurally just adjudication mechanisms, they were less likely to repeat the violations. Therefore, as Tyler et al. concluded, the users were more likely to self-regulate and follow the top-down rules when Facebook exercised procedural justice in its dispute resolution process. These findings speak to the first issue noted: the desirability of self-regulation in response to viewing the platform as a legitimate authority.

A further goal not addressed in the Facebook study is the ability of these same fair procedures to enhance the online climate on a site. We will discuss this goal in the next section in more detail.

---

[61] *About Moderation,* NEXTDOOR, https://help.nextdoor.com/s/article/About-moderation? (last visited Jan. 23, 2021).

[62] *About Community Reviewers and Moderation,* NEXTDOOR, https://help.nextdoor.com/s/article/Community-Reviewers-and-Moderation?.

The goal of many platforms is to create a safe climate within which people can constructively discuss emotional and potentially divisive issues in their lives and communities. People's ability to do so is also affected by whether they trust the authorities creating and managing the platform through which they are interacting. Again, legitimacy is key to providing a baseline level of comfort and reassurance that can enable such dialogue.

Promoting Prosocial Behavior

Prosocial approaches are different from simply trying to avoid harms or violations, no matter whether one is focused on criminal justice system outcomes or trying to ensure compliance with content moderation rules online. Prosocial approaches treat a positive social environment as a goal.[63] If the platform and its users create a positive social environment, the need for control by the platform is reduced because that social environment produces socially desirable outcomes.

While procedural justice—based approaches can enhance rule-following by motivating voluntary compliance, we think prosocial approaches based on procedural justice theories can do more: platforms can use them to motivate users to do good. To promote prosocial behavior, platforms must increase community engagement, individuals' desire to pursue a collective goal, and engage in economic and political activities. Engagement is involvement with one's own community.[64] Specifically, it is discretionary cooperation, meaning that instead of just following the rules authorities impose and cooperating with the authority, the community proactively behaves in such a way that the members

---

[63] For a deeper understanding of prosocial approaches in criminal justice system and online platforms, refer to the website of Justice Collaboratory based in Yale Law School: https://law.yale.edu/justice-collaboratory.
[64] Tyler & Jackson, *supra* note 57.

trust one another and know that if a problem arises, they can face it collectively.[65]

For a community to engage (offline or online), the individuals must identify with the values of the community and be willing to act on behalf of the collective. The decision-makers and authorities can incentivize the community to engage with one another by being legitimate. Engagement can increase when the community members have normative alignment with one another and identify with the values and goals of the community.[66]

One approach to increasing engagement is to create virtual social structures. According to prosocial theories, social structures can create opportunities for communities to thrive and cooperate.[67] These structures in real life are gyms, town halls, youth centers, bars, bistros, and the like. These social structures are the heart of community vitality in the real world.[68]

We can translate social structures to their online analogues. For example, online forums, town halls, and groups, and even some algorithms and other virtual tools, can play a role in building a strong social structure. Black et al. also mentioned that even simpler communication systems such as email lists can help in providing social structures. Byrne argued that the virtual "forums" on websites are where community vitality is happening and people engage. He called these forums central to public life and an opportunity to

---

[65] *Id*. at 81.
[66] *Id*. at 84.
[67] As Meares argued, where social structures are weak, it is difficult to exert social control. Thus, to be able to govern online communities through self-regulation and social control, it is necessary to provide the social structures. Meares, *supra* note 55.
[68] RAY OLDENBURG, THE GREAT GOOD PLACE: CAFES, COFFEE SHOPS, BOOKSTORES, BARS, HAIR SALONS, AND OTHER HANGOUTS AT THE HEART OF A COMMUNITY (1999).

understand how various communities construct, modify, and stabilize.[69]

Historically, community vitality was generated through instant messengers, chat rooms, weblogs, and discussion boards. For example, chat rooms became the social structures where users could discuss the rules and responsibilities governing their behavior in their online community. The effect of cyberspace on physical world communities, not to mention the fact of online communities that depend on cyberspace for existence, has been profound. The effect even inspired predictions that as bars, restaurants, and other places came to lose their sense of community vitality, perhaps online communities would replace them and bring community vitality.[70]

To advance prosocial interaction, platforms must enhance engagement with political, social, and economic activities. Tyler and Jackson identified the following indicators of engagement (actions to help the community and its vitality):[71]

- Perceived social capital (community members helping each other and working together to bring safety)
- Community identification (being proud of your community)
- Political capital (engaging with changing political decisions)
- Economic activities (going to shops and restaurants and spending time with the community)

Tyler and Jackson argued further that procedural justice is associated with indicators of engagement. We can theorize that if procedural justice criteria are satisfied in an online group, it is likely that engagement also will increase. The theory's hypothesis is that

---

[69] Byrne, *supra* note 30.
[70] OLDENBURG, *supra* note 68.
[71] Tyler & Jackson, *supra* note 57, at 79.

if people are treated well (fairly, with respect and dignity) by those they encounter in a given community, they are more likely to engage with voluntary actions, build social capital, and get involved with economic activities. In the next section, we describe how to measure a prosocial production system. That way, the measures can feed back to generate the desired behavior and meet the criteria stated above.

Improved Measurements for a Prosocial Production System

How should platforms create an environment in which prosocial activity begets more prosocial activity, creating a positive feedback loop that ensures a good online social environment? In other words, how should platforms set up a prosocial production system? The first step for the platforms is to select a prosocial goal for themselves. The goal could be to achieve healthy interaction or enhance civility. In order to operationalize "healthy interaction" or "civility," we define them and determine the constitutive elements. For example, we can operationalize civility by asking the extent to which a candidate action exhibits tolerance and respect. It is also important to have an understanding of what constitutes tolerance and respect and how to measure the increase or decrease of each. Using legal and social science methods, we can discover the constitutive elements of respect and tolerance.[72]

---

[72] Scholars across various disciplines have discussed how to define and operationalize prosocial goals such as civility and healthy online interactions. *See* Jeremy Waldron, *Civility and Formality*, NYU SCHOOL OF LAW, PUBLIC LAW RESEARCH PAPER NO. 13-57 (2013); Zizi Papacharissi, *The Virtual Sphere: The Internet as a Public Sphere*, 4 NEW MEDIA & SOC'Y 9 (2002); Arthur Santana, *Virtuous or Vitriolic: The Effect of Anonymity On Civility in Online Newspaper Reader Comment Boards*, 8 JOURNALISM PRACTICE 18 (2014); Myiah Hutchens et al., *What's in a Username? Civility, Group Identification, and Norms*, 16 J. INFO. TECH. & POL. 203 (2019). Chris Vargo & Toby Hopp, *Socioeconomic Status, Social Capital, and Partisan Polarity as Predictors of Political Incivility on Twitter: A Congressional District-Level Analysis,* 35 SOC. SCI. COMPUT. REV. 10 (2017).

The next step is to set up the virtual social structures for a sample of individuals. For example, as discussions are heating up and are becoming controversial on some general thread, the platform can empower the poster by recommending the creation of a group. There must also be policies and methods that encourage people to do good—for example, prompts that would pop up in the form of pithy messages when users join a group are having a conversation.

Finally, it is critical to measure these efforts. Platforms are run (perhaps even "overrun") with attention to metrics. If they cannot measure it, they will not do it. We have identified the need for "measurement" in our conversations with platforms when discussing strategies for enhancing healthful interactions. Using metrics and measurements is a good way to improve decision-making processes; however, the platforms need to enhance and modify their approach and update their metrics. Thus, in this paper, we also provide some suggestions and benchmarks for measuring social phenomena, with the hope to improve and standardize measurement benchmarks on platforms. [73]

In collaboration with one platform, we have undertaken a study that implements some of these suggestions. For example, we have designed prompts and messages based on the procedural justice indicators. These prompts are displayed when the users enter a virtual social structure such as a group. The prompts encourage and remind the community members of the platform's guidelines and ask the community members to "listen to each other" (allowing for participation), "lead with compassion" (respect others with

---

[73] Other scholars have also come up with methods to measure prosocial behavior online. For example, see Jiajun Bao et al., *Conversations Gone Alright: Quantifying and Predicting Prosocial Outcomes in Online Conversations*, *in* PROCEEDINGS OF THE WEB CONFERENCE (2021). Bao et al. have created a process through which we can quantify prosocial outcomes on platforms.

dignity), "cite sources" (maintain neutrality and be objective), and "take other people's issues seriously" (show good faith).

Measuring how much the prosocial governance mechanism produces prosocial behavior can help reform the governance mechanisms based on science and not clairvoyance. To measure prosocial compliance, which means feeling obligated and motivated to follow rules, we need a less outcome-oriented approach than the one usually followed by platforms. Prosocial compliance does not only mean that people comply with the rules, but also that people self-regulate and do not turn into repeat offenders.

One solution for a less outcome-oriented approach is to use community as the unit of analysis. Thus, instead of measuring only relations between or among individuals, we should measure individuals' relations with the community. To measure whether people have internalized rule-following, as Tyler et al. have previously done,[74] we can use survey strategies to measure whether providing virtual social structures and treating the users with procedural justice has had any effect on following the rules. The surveys should also ask why the members followed the rules, out of self-interest or out of norm alignment with the community, and the perceived legitimacy of the decision-makers.

We can measure community engagement through the indicators mentioned in the previous section: do they volunteer to help their online community members, are they proud of the online community they belong to, have they accumulated social or political capital and engaged more with economic activities? Through a survey, people can indicate how likely they are to attend online political activities on the platform, get engaged with transactions, or

---

[74] Tyler et al., *supra* note 2.

intervene if they see members being disrespectful to each other, as well as whether they are proud or feel good about being involved with an online group.[75]

This method is, however, insufficient, as it measures the users' opinions *post-factum* and is not an observation of actual behavior. There are other methods that can measure prosocial behavior during ongoing interactions. For example, an indicator for cooperation is community-led efforts to inform the authorities of a problem. We can also control for the increase or decrease in the number of voluntary initiatives that community members come up with in order to help the decision-makers and the community leaders bring more civility to the platform or increase healthy interactions.

An additional way is to control for changes in prosocial indicators by observing the communities' social, political, and economic activities on the platform. For example, we can consider an increase or decrease in participation in voting and creating sub-communities to discuss politics. We can also measure the increase or decrease in participation in collective actions (for example, an online fundraising event). If the platform is multisided, i.e., it facilitates transactions as well as interactions, engagement can also be measured by controlling for an increase or decrease in economic activities (the rate of buying and selling on the platform).

A common approach to measuring the effect of governance mechanisms that communications and human—computer interaction scholars use is sentiment and textual analysis. Scholars use sentiment analysis and textual analysis to measure offensive

---

[75] We mentioned the criteria in *supra* Section II(B). As a reminder the criteria are: Community identification (being proud of your community), Perceived social capital (community members helping each other and working together to bring safety), Political capital (engaging with changing political decisions), and Economic activities (going to shops and restaurants and spending time with the community).

words or hate speech in a certain corpus of text (in the case of a platform, some set of messages on it).[76] The software often works based on a lexicon that can assess the tone of the text and label it as positive, negative, or neutral.[77] Software often has a training component so that the software can be tailored within some limits to the likely normal baseline of sentiment found in an average text from a given source. The trained software can measure the rate of positive, negative, and neutral words based on the number of occurrences and provide an estimate of how negative or positive certain texts are. The positive and negative sentiments can be correlated with various prosocial values—for example, the positive sentiment can be civil interactions, and the negative can be uncivil interactions. However, it is critical to first train the software with what is perceived as civil or uncivil to attain better results

CONCLUSION

Over time, as platforms became both commercialized and centralized, their approach to governance changed. Instead of fostering the communities that existed on their platform, they used a top-down deterrence-based mechanism to govern their platforms. This meant that they did not work on creating tools for empowering these communities, but tools to govern their users (primarily on an individual basis) and content. Compliance in these platforms does not mean motivating users to comply with the rules and regulations. Rather, platforms force their users to comply through deterrence-

---

[76] There are many studies that use this method, using different software and hate-speech or offensive speech lexicons. One example is Rishab Nithyanand et al., *Measuring Offensive Speech in Online Political Discourse*, *in* 7TH USENIX WORKSHOP ON FREE AND OPEN COMMUNICATIONS ON THE INTERNET (2017).

[77] Papacharissi, *supra* note 72; Federico Neri et al., *Sentiment Analysis on Social Media*, *in* Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining 919 (2012).

based mechanisms such as removal of content, suspension, and blocking.

We believe one way to reform the governance mechanisms of these platforms is by designing and implementing a prosocial production system. In this paper, we have presented a self-motivated prosocial production system to reform platforms' punitive approach to governance, successfully limit negative behavior, and promote positive behavior. We have conceptualized the process through which we apply the theory of procedural justice to platforms' governance mechanisms. We have also laid out the steps for designing a prosocial production system. Finally, we have presented a system through which various elements necessary for community vitality and prosocial behavior can be measured.