

Haggart, Blayne; Iglesias Keller, Clara

Article — Published Version

Democratic legitimacy in global platform governance

Telecommunications Policy

Provided in Cooperation with:

WZB Berlin Social Science Center

Suggested Citation: Haggart, Blayne; Iglesias Keller, Clara (2021) : Democratic legitimacy in global platform governance, Telecommunications Policy, ISSN 0308-5961, Elsevier, Amsterdam, Vol. 45, Iss. 6, pp. --,
<https://doi.org/10.1016/j.telpol.2021.102152>

This Version is available at:

<https://hdl.handle.net/10419/238195>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

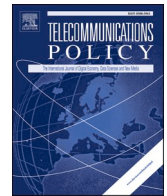
Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>



Democratic legitimacy in global platform governance

Blayne Haggart^{a,*}, Clara Iglesias Keller^b

^a Brock University, 1812 Sir Isaac Brock Way, St. Catharines, Ontario, L2R 4E4, Canada

^b WZB Berlin Social Science Center, Reichpietschufer 50, D, 10785, Berlin, Germany

ARTICLE INFO

Keywords:

Platform governance
Democratic legitimacy
Human rights
Input legitimacy
Global governance

ABSTRACT

The goal of this paper is to propose a democratic legitimacy framework for evaluating platform-governance proposals, and in doing so clarify terms of debate in this area, allowing for more nuanced policy assessments. It applies a democratic legitimacy framework originally created to assess the European Union's democratic bona fides – Vivian Schmidt's (2013) modification of Scharpf's (1999) well-known taxonomy of forms of democratic legitimacy – to various representative platform governance proposals and policies. The first section discusses briefly the issue of legitimacy in internet and platform governance, while the second outlines our analytical framework. The second section describes the three forms of legitimacy that, according to this framework, are necessary for democratic legitimation: input, throughput and output legitimacy. The third section demonstrates our framework's utility by applying it to four paradigmatic proposals/regimes: Facebook's Oversight Board (self-governance regimes); adjudication-focused proposals such as the *Manila Principles for Intermediary Liability* (rule-of-law-focused regimes); the human-rights-focused framework proposed by then-UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression; and the United Kingdom's Online Harms White Paper (domestic regime). Section four describes our four main findings regarding the case studies: non-state proposals seem to focus on throughput legitimacy; input legitimacy requirements are frequently under examined; state regulation is usually side-lined as a policy option; and output legitimacy is a limited standard to be adopted in supranational contexts. We conclude that only by considering legitimacy as a multifaceted phenomenon based in democratic accountability will it be possible to design platform-governance models that will not only stand the test of time, but will also be accepted by the people whose lives they affect.

1. Introduction

The issue of legitimacy looms large over the question of how best to regulate global online platforms. Most popular discussions of Facebook's Oversight Board (Clegg, 2020), for example, have highlighted the company's hope that it will bring "legitimacy" to its operations and its preference for self-regulation (e.g., Douek, 2019). Recent academic work, particularly that by Nicholas Suzor, has similarly focused on questions of legitimacy in platform governance (e.g., Suzor, Geelen and West, 2018; Suzor, 2019 and 2018; see also Bloch-Wehba, 2019; Gorwa, 2019; Gorwa, 2019a; Kaye, 2019), and can be seen as a continuation of a long-standing debate around legitimacy in internet governance. In this realm, the quest for justifications for the internet's private ordering is a common theme, from technical protocol design by "private standards-setting institutions or individual companies" (DeNardis, 2012, 723) to the policy

* Corresponding author.

E-mail addresses: bhaggart@brocku.ca (B. Haggart), clara.keller@wzb.eu (C.I. Keller).

decisions regarding speech made or enforced by content intermediaries (DeNardis 2012, 725; see also Johnson and Post, 1996; Reidenberg, 1996; Lessig, 2006). More recently, proposals intending to encourage greater legitimacy in platform regulation have appealed to rule-of law principles (Suzor, 2019; Manila Principles, 2015), as well as to human rights laws and norms (Kaye, 2019).

This article argues that while appeals to ground platform regulation in human-rights laws and the rule of law are commendable, they do not offer a solid foundation for ensuring the legitimacy of a platform-governance regulatory framework, and thus to its continued success. In particular, we argue that insufficient attention has been paid to the issue of democratic legitimacy in global-platform governance. Rather than treat legitimacy as a one-dimensional concept (based, for example, on the presence of transparency in rule-making, adherence to free-speech norms, or the rule of law), it argues that legitimacy is a multifaceted phenomenon that must be assessed as such. We propose a three-dimensional framework for assessing the democratic legitimacy of platform governance proposals and regimes that draws on Vivian Schmidt's (2013) modification of Scharpf's (1999) well-known taxonomy of democratic legitimacy as it relates to the European Union. According to this framework, the democratic legitimacy of a policy regime can be divided into three parts: input legitimacy, throughput legitimacy, and output legitimacy. Although conceptually separate, these different forms of legitimacy interact in ways that may reinforce or undermine each other. Proposals that focus on one form of legitimacy at the expense of the others can have the unintended consequence of actually undermining overall legitimacy.

Applying this framework to platform governance frameworks and proposals allows us to highlight their legitimacy-related strengths and weaknesses in a more nuanced way. It highlights how different aspects of policy design contribute in differing, sometimes unequal ways, to the construction of legitimacy. In differentiating among various facets of legitimacy it also offers a clearer conceptual framework for developing and analyzing future proposals. In particular, it highlights the extent to which proposals that are not rooted in domestic-democratic legitimacy (i.e., proposals for regulation outside of the state) tend to overemphasize throughput legitimacy to the detriment of input and output legitimacy. Within the context of the complex nature of the internet's normative order (Kettman, 2020), they also tend to sideline the possibility of state regulation (including by democratic states) – which faces its own difficulties of legitimate and efficient law enforcement in digitalized spaces. Possibly due to this reluctance to consider state regulation, platform governance initiatives show a tendency to take the for-profit, global nature of these platforms as an unchangeable fact of nature. Consequently, the legitimacy of these non-state-based regimes are likely to be much more precarious than their proponents may realize.

This paper is organized as follows. The first section discusses briefly the issue of legitimacy in internet and platform governance, while the second outlines our analytical framework. The third section demonstrates our framework's utility by applying it to four paradigmatic proposals/regimes: Facebook's Oversight Board (self-governance regimes); adjudication-focused proposals such as the Manila Principles for Intermediary Liability (rule-of-law-focused regimes); the human-rights-focused framework proposed by the then-UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, David Kaye; and the United Kingdom's Online Harms White Paper (domestic regimes). The fourth section highlights four main findings regarding the case studies: non-state proposals seem to focus on throughput legitimacy; input legitimacy requirements are frequently under examined; state regulation is usually side-lined as a policy option; and output legitimacy faces important limitations regarding supranational regulation in a pluralist world. The paper concludes with some final thoughts about the implications of our analysis for future policymaking. Ultimately, we argue that only by considering the multifaceted nature of legitimacy, based in democratic accountability, will it be possible to design platform-governance models that will not only stand the test of time, but will also be accepted by the people whose lives they affect.

2. Democratic legitimacy in internet and platform governance

2.1. Defining legitimacy

We argue that democratic legitimacy is the appropriate starting point for assessing the legitimacy of platform governance regimes and proposals. Communication and communication infrastructure play a “unique public role” in society, shaping culture, economies and societal relationships. Historically, this has always led policymakers to implement “public-regarding models” of governance legitimized by public processes, instead of subject to self-regulatory organizations (Sylvain, 2010, pp. 209–210).

In this spirit, legitimacy in platform governance raises concerns due to the long-diagnosed democratic deficit to the virtual space's private ordering. This democratic deficit refers to the fact that private companies make the choices that set norms and directly influence the behavior of billions of users (Reidenberg, 1996; Lessig, 2006). This strong corporate influence on rule- and norm-setting raises concerns about the interests behind these choices, and the processes that led to them and their binding nature. As we address below, the platform governance-specialized literature has addressed these concerns in different manners and, as we will argue, have failed to do so in a comprehensive way. This analysis will be informed by some of the concepts that have long supported the justification of state's dominance over individuals, and that to a different degree, have also inspired the seek for legitimacy outside of the state.

Within the nation-state, a policy or law's legitimacy is associated with different elements that revolve around the question of who makes the laws that apply to everyone. For this reason, citizenship participation plays a pivotal role, and legitimacy can be often understood as the “realization of democracy through popular sovereignty” (von Bogdandy, 2004, p. 887). This notion is centered on one's “right to take part in the government of his country” and on “the will of the people [which] shall be the basis of the authority of the government” (UDHR Article 21). If approached through the individual's perspective, legitimacy can also be understood as internal justification; i.e., as a “moral duty to obey the law” (Kumm, 2004, p. 908) – together with “the willingness to obey laws even when these go against their own interests” (Schmidt, 2013, pp. 9–10; Giddens, 1984, cited in Epstein 2013). Legality, meanwhile, is also

associated with democratic legitimacy as it refers to “domination (...) by virtue of the belief in validity of legal statute and functional competence based on rationally created rules” (Weber, 1919). The requirement of covenant pre-established rules for the political regime to be built upon can also be expressed through the *rule of law* (as formalization of means to “limit broad discretionary powers of public regulators by ensuring that decisions are lawfully made” (Suzor, 2019, p. 106))) and *constitutionalism* (which presumes “a legal statute, providing for equal fundamental and political rights and their effective protection, transparency, accountability and the respect of the rule of law” (Pernice, 2015, p. 27)).

These different arguments for how democracies legitimize themselves – popular participation, individual constrictions, the rule of law and constitutionalism – also guided the quest for justification outside of the nation-state, which becomes more relevant as globalization and multiple privately defined orders expand. Even though the lack of electoral representation in such spaces turns this into a rather complex task, legitimacy here can also be driven by different expressions of the same concerns.

The debate over the legitimacy of international law offers a good illustration of the difficulty of assessing legitimacy for rule-making outside of the democratic nation-state. International law is a framework that exerts influence at a collective level, but is not governed by a constitution or elected representatives. In this realm, individual constrictions is the legitimacy criterion adopted, for instance, by Kumm (2004). According to Kumm, the legitimacy challenge refers to the question of to “what extent should citizens regard themselves as morally constrained by international law” (Kumm, 2004, p. 908). Taking a different approach, Bogdandy associates the legitimacy of international law with popular participation (2004). He focuses on the school of thought according to which state supremacy is assumed to continue over other spheres of society, which would presume the expansion of democratic legitimacy to other spaces. This presumption could rest either on a unitary model of legitimacy (according to which public acts achieve a democratic quality only when they can be traced back to the chain of democratic legitimation), or on a pluralist one (where the implementation of new forms of civic participation would support the legitimacy claim of globalized frameworks) (Bogdandy, 2004, p. 903).

Overall, the assurance of legitimacy of international law cannot be directly transposed to the platform-governance debate. International law is in many cases performed by public actors, and for this reason its legitimacy can still be ultimately derived from democratic representation within democratic nation-states (as suggested by Bogdandy’s unitarian model of legitimation). That said, the governance of globe-spanning platforms also involves an exercise in justifying broadly applicable rules that are not embedded within processes of electoral representation – an exercise that has been ongoing since the internet expanded to civil use and millions of users at a global level became bound to its inescapable private ordering.

2.2. The problem of democratic legitimacy in internet and in platform governance

Questions about who has legitimate authority to regulate the internet, as well as what constitutes legitimacy in internet governance, are almost as old as internet governance itself. Since the digital sphere presents itself as an environment essentially influenced by private regulation, early literature labelled as “exceptionalist” argued the irrelevance of traditional measures of democratic legitimacy. Johnson and Post (1996), for example, argued against government control of the internet and in favor of the need for a new system, free of state-centered notions of jurisdiction. While these arguments have been challenged by those who have noted that the internet could not be immune “to the longer and deeper forces that shape human history” (Wu, 2010, p. 180), the challenge of finding a source of legitimacy outside the reach of majoritarian institutions of electoral representation (Scharpf, 1996; Schmidt, 2013) has remained relevant.

Multistakeholderism as a governing philosophy has been widely adopted, particularly with respect to standards and protocols governance, with multistakeholder processes serving “primarily as venues for dialogue and coordination” (Pohle, 2016, p. 5). Multistakeholderism as a governance approach is rooted in the idea that the governance process is best served by including all interested stakeholders as well as those with relevant expertise. It has been criticized on democratic-accountability grounds for failing to involve the wider citizenry in deliberation, raising questions about “how the public interest is reflected in protocol design and how, procedurally, these entities derive the necessary legitimacy to make such design decisions” (DeNardis, 2012, p. 723; see also Sylvain, 2010).

As platform governance emerged as a key area of internet governance, internet governance scholarship has moved from a “code is law” focus to further exploring the different aspects of intermediaries’ private ordering. As a result of platforms’ “private mediation between Internet content and the humans who provide and access this content” (DeNardis, 2012, p. 725), platform-governance legitimacy issues are intertwined with discussions about the scope of freedom of expression, access to information and other fundamental rights (Kaye, 2019). Social networks and search engines, for instance, play a by-now-notorious role in matters such as privacy and online advertising (Gasser, 2006), censorship requests from governments (DeNardis, 2015), and regulation of speech in a broad sense.

For this reason, legitimacy-related platform-governance concerns tend to revolve around the fact that “decisions about what we can do and say online being made behind closed doors by private companies is the opposite of what we expect of legitimate decision-making in a democratic society” (Suzor, 2019, p. 8). These include, for instance, the limits of content moderation policies (Gillespie, 2017), both led by workers (Roberts, 2019) and automated tools (Gorwa, Binns, & Katzenbach, 2020). Recently, literature also highlights how platforms’ private decision-making involves “complex processes through which social networks develop norms, who they involve in these processes and – importantly – how actors within these companies conceive of their norm-setting function” (Kettmann and Schulz, 2020).

Different authors have proposed different approaches to the problem of assessing and improving the legitimacy of global online platforms. These proposals usually focus on specific aspects of platforms’ decision-making processes, such as greater user participation, human-rights standards or procedural guarantees. This narrow focus, however, narrows their potential for providing a robust, complete framework for assessing legitimacy in platform governance.

In an attempt to bring an approximation of democratic participation to platform rule-making, Ivar Hartmann, among others, advocates for greater user engagement in content moderation, which would fulfill what they see as “Internet users’ deep-rooted wish for self-governance” (2017, p. 22). This is in line with internet exceptionalism proponents (Barlow, 1996) who advocated for an internet ruled by its users instead of the traditional state-centered power structures. Although initially prominent in the discourse and the literature, this form of democratization has come to be seen with a large degree of scepticism. No matter what mechanisms are employed to assure user participation, this could still not be equated to the expansion of the legitimate power of citizens. Echoing Carr (2015) and Powers and Jablonski’s (2015) criticism of how multistakeholderism reinforces existing power dynamics, Stivers (2018, p. 154) rejects user-participation centered proposals as a means for replicating traditional forms of democratic deliberation, arguing, “we cannot safely take it for granted that the “collaborative governance” that occurs in networks and platforms is the same as the kind that takes place in physical public space.” There would be no means to assure participation consistent with institutional democratic mechanisms in a decision-making process that would still be inevitably stirred by private actors.

The adoption of human rights has also been proposed as a legitimacy standard (Kaye, 2019), alongside transparency in platform decision-making and other due process related guarantees (e.g., Suzor, Van Geelen and Myers West, 2018). Suzor, Van, and West (2018) is notable for its direct engagement with the question of determining the legitimacy of platform governance, linked inextricably to the extent that regulations follow “human rights values” and “procedural values”: “the rule of law, due process and transparency, as well as “participation in decision making” (Suzor, Geelen, and West 2018, p. 387; pp. 391–92). The idea of using a “set of procedural limits on how rules are made and enforced” to promote governance legitimacy is also developed in Suzor (2019, p. 115). Overall, these arguments are close to the concept of digital constitutionalism, which refers to “a constellation of initiatives that seek to articulate a set of political rights, governance norms, and limitations on the exercise of power on the Internet” (Redeker, Gill, & Gasser, 2015, p. 303), in order to ground internet governance in “fundamental rights and democratic principles” (Padovani and Santaniello, 2018, p. 296; see also Redeker, Gill & Gasser, 2015; Celeste, 2018; and De Gregorio, 2020).

Although they provide a comprehensive list of high-level human-rights values to which platforms should adhere in order to legitimize their actions and rule-making (Suzor, Geelen, and West, 2018, pp. 388–89), the analysis by Suzor et al. fails to come to terms with the fundamental question of how these rights should be fulfilled, and that good-faith actors may interpret these high-level rights in radically different ways. Consider the operationalization of a concept such as “freedom of expression and privacy rights” (Suzor, Geelen, and West, 2018, p. 390). The promotion of individual privacy rights, for example, may conflict with other legitimate, rights-based objectives. In British Columbia, Canada, for instance, the provincial government in September 2020 refused to provide First Nations communities with information these communities said was necessary to fight the COVID-19 pandemic, “on the grounds that it could violate the privacy of patients and put some of them at risk of facing discrimination” (Xu & Hager, 2020).

Operationalizing freedom of expression is similarly fraught. It is increasingly understood that the voices of marginalized individuals and groups can be drowned out online through a torrent of abuse and misinformation, actions that are not traditionally seen as censorship but that can have the same effect on the ability to express oneself freely (Tufekci, 2018). Furthermore, appeals to high-level human-rights frameworks do not fully consider the reality that “freedom of expression” can be and is interpreted in different ways, even among and within democratic societies such as those in Europe and the United States (Franks 2018). Ultimately, the extent to which a country or company adheres to “human rights values” will be in the eye of the beholder.

As well, while questions of values and process as legitimizing forces are discussed extensively, the issue of which actors possess rule-setting legitimacy here receives short shrift. Part of this lack of attention to the decision-making question reflects the reality that these issues are largely ignored in platform-governance human rights indexes or evaluation projects; in Suzor et al.’s survey of such projects, only Ranking Digital Rights (RDR) addresses this issue. Even the RDR criteria only considers the extent to which a company engages in multistakeholder governance, which – as noted above – is a weak proxy for the rights set out under the Universal Declaration of Human Rights’ (UDHR) Article 21 (Suzor, Geelen, and West, 2018, p. 392). Most importantly, by taking global platforms’ current private-ordering regime as a given rather than something that can be changed through regulation, their analysis leaves no room to consider even a theoretical role for the state as regulator.

To varying degrees, such analyses and proposals have attempted to verify whether digital norms and structures are steered in the direction of the public interest. This brief assessment of the literature on internet and platform governance suggests the need for a multidimensional approach to assessing governance in these areas. Analyses of governance of the internet’s infrastructure layer suggests that expertise and multistakeholderism do not take us far when it comes to ensuring representativeness; user participation cannot be equated to legitimate participation; and human rights-based frameworks are limited by the lack of a global consensus on the content of these principles as opposed to the general sentiment behind them. At the heart of the issue is that, in situations in which there are good-faith disagreements about which rules should be implemented, the process of how these decisions are made, and by whom, takes on a supreme importance. While the aforementioned approaches are useful for thinking about particular aspects of platform governance, they fail to offer a convincing way either to assess or to provide a response to the challenge of platform-governance legitimacy, particularly as it relates to democratic legitimacy. In the next section, we outline a more nuanced theoretical framework that can more comprehensively assess platform governance legitimacy claims.

3. Assessing democratic legitimacy outside the nation-state

Assessing democratic legitimacy within a state is relatively straightforward. In a liberal-democratic state, universal suffrage, exercised on a regular basis, provides a government with the legitimacy to pass laws and enact policies under which citizens live. Although governments are expected to act within socially acceptable norms, which itself confers a form of legitimacy on the government, the ballot box represents the ultimate sanction on rulers, who are judged based on their words and deeds while in power.

Power and legitimacy (again, in theory, and generally speaking) flow from the citizenry, through the ballot box, to the government and the state.

For European political theorists, the European Community and European Union presented a challenge to the question of the origins of democratic legitimacy. As a supranational entity made up of inter-state representation, supranational institutions and only a relatively weak representative body, the EU cannot claim legitimacy in the same way as the democratic nation-state.

As Vivien A. Schmidt (2013) notes, the discussion regarding EU democratic legitimation has long been framed around Fritz Scharpf's (1970) typology of input and output legitimacy. Input legitimacy refers to the "EU's responsiveness to citizen concerns as a result of participation by the people," while output legitimacy refers to the "effectiveness of the EU's policy outcomes for the people," input legitimacy refers to the "EU's responsiveness to citizen concerns as a result of participation by the people" (Schmidt, 2013, p. 2). To this, Schmidt adds a third category, "throughput legitimacy," which highlights the quality of the governance process and "is judged in terms of the efficacy, accountability and transparency of the EU's governance processes along with their inclusiveness and openness to consultation with the people" (Schmidt, 2013, p. 2).

The European Union's situation, as a collection of nation-states and institutions in an overarching supranational entity, presents a helpful analogy to those of global American platforms such as Google and Facebook. Like the European Union, these platforms may be seen as consequential supranational rule-setters with a complicated relationship with the nation-states in which they operate. Although in practice platforms are run by one or a small group of businesspeople, the multistakeholder ideal of internet governance can potentially involve many groups in platform governance. Furthermore, as they operate globally, their rules can affect societies that hold different values. Perhaps most significantly for our purposes, platforms, in their current, for-profit form, cannot ground their legitimacy in direct accountability to the citizenry – even though their privately designed policies still affect billions of users. This situation has something of a parallel in the European Union, for which the existence of a "democratic deficit" in terms of (weak) electoral accountability has been a longstanding concern (see discussion in Murdoch, Connolly and Kassim 2018, p. 389). Similar, albeit more severely, than the EU, digital platforms cannot, in their current, for-profit form, base their legitimacy in accountability to a citizenry. These similarities, we argue, make Schmidt's three-pronged framework a useful way to assess legitimacy in these complex and hybrid digital communication spaces.

Although we argue that this comparison is useful for thinking about questions of legitimacy, it should not be taken too far. Unlike the European Union, which was the creation of numerous states, online platforms are primarily headquartered in one state, the United States, which exercises disproportionate influence over them, and whose norms remain embedded in these platforms (Klonick, 2017, p. 1599). Second, while some observers may be dazzled by the size and apparent influence of these platforms, they remain corporations, not governments. Their relationship to actual states is therefore different from that of the EU to its member states.

In the remainder of this section, we expand on the meaning of the three types of legitimacy and detail what this evaluative framework entails.

3.1. Input legitimacy

Input legitimacy "refers to the participatory quality of the process leading to laws and rules as ensured by the 'majoritarian' institutions of electoral representation" or the institutions' "responsiveness to citizen concerns as a result of participation by the people." It is grounded in the citizenry and draws its legitimacy from claims to represent the citizenry. It includes not only the presence or absence of majoritarian (i.e., democratic) institutions and elections, but also on the presence or construction "of a sense of collective identity and/or the formation of a collective political will" (Schmidt, 2013, p. 5). It "depends on citizens expressing demands institutionally and deliberatively through representative politics while providing constructive support via their sense of identity and community" (Schmidt, 2013, p. 7). It also, Schmidt notes following Scharpf (1999), requires "thick collective identity and a European [in the case of the European Union] demos" (Schmidt, 2013, p. 4).

Input legitimacy therefore focuses on the questions, "Who is the citizenry, and what is the 'participatory quality'" (Schmidt, 2013, p. 4) of their involvement in governing? This focus on input legitimacy is important because it forces platform-governance scholars and advocates to confront head-on the issue of who governs. Input legitimacy requires some form of democratic representation based in a citizenry, which itself has to have some form of collective identity. Input legitimacy does not always directly require that policy be made by majoritarian institutions – all democratic countries contain their fair share of non-democratic institutions that prioritize technical proficiency. However, these institutions, such as central banks or arm's length regulators, possess legitimacy not just because they possess a technical expertise or are isolated from politics. Rather, they can claim legitimacy "because they operate in the 'shadow of politics', as the product of political institutions, with political actors who have the capacity not only to create them but also to alter them and their decisions if they so choose" (Schmidt, 2013, p. 10).

3.2. Throughput legitimacy

Throughput legitimacy "focuses on the quality of ... governance processes." It is "process-oriented, and based on the interactions – institutional and constructive – of all actors engaged in ... governance" (Schmidt, 2013, p. 5). It "involves first of all the efficacy of the many different forms of ... governance processes and the adequacy of the rules they follow in policy making" (Schmidt, 2013, p. 6). It "demands institutional and constructive governance processes that work with efficacy, accountability, transparency, inclusiveness and openness" (Schmidt, 2013, pp. 7-8).

The mechanisms of throughput legitimacy, meanwhile, are "their efficacy, accountability, transparency, inclusiveness and openness to interest intermediation" (Schmidt, 2013, p. 6). Efficacy refers to how smoothly the processes of rule-making and

implementation function. Accountability involves the extent to which “actors are judged on their responsiveness to participatory input demands and can be held responsible for their output decisions” (Schmidt, 2013, p. 6).

Transparency, meanwhile, refers to whether “citizens have access to information about the process and [whether] decisions as well as decision-making processes in formal ... institutions are public” (Schmidt, 2013, p. 6). Transparency should not be confused with accountability, which “demands some form of scrutiny by a specific forum” (Schmidt, 2013, p. 6) and is thus related to input legitimacy. Rather, transparency is merely “a prerequisite to accountability” (Schmidt, 2013, p. 6). Finally, inclusiveness and openness of institutional processes to civil society participation in the policymaking process is legitimizing as a balance to “access and influence among organized interests representing business” (Schmidt, 2013, pp. 6–7).

Since the major platforms are themselves businesses, in the context of private platform governance civil-society groups (which would also include private-sector actors affected by the platform in question) will necessarily find themselves balancing against the platform’s private, for-profit interests. Throughput participation is distinctive from the citizen participation covered by input legitimacy because it involves the integration of “interest articulation” in the governance process. In contrast, input legitimacy involves attempts to influence (or select) “elected representatives” (Schmidt, 2013, p. 7). This distinction is particularly useful when considering calls for greater multistakeholder participation in platform governance. To the extent that they are involved in the process of interest articulation, multistakeholder involvement in platform regulation, such as the type highlighted in 4.3.2 below, should be seen primarily as part of the throughput process of governance.

3.3. Output legitimacy

Output legitimacy, meanwhile, covers “the problem-solving quality of the laws and rules” (Schmidt, 2013, p. 4), or “effectiveness of ... policy outcomes for the people” (Schmidt, 2013, p. 2). In considering output legitimacy, EU democratic-legitimacy scholars have considered a range of issues, including “the community enhancing performance of the policies themselves,” as well as, from a constructivist perspective, the extent to which these policies are communicatively legitimated through discursive and communicative actions (Schmidt, 2013, p. 5).

It is not, however, enough for policies to solve a problem effectively: policies must “work effectively while resonating with citizens’ values and identities” (Schmidt, 2013, p. 7). Here, we see how output legitimacy is not just about getting the “correct” outcome, but the correct outcome grounded in citizens’ perceptions of what the correct outcome is. As economist Dani Rodrik points out with respect to financial regulation, everybody wants prudent and safe financial regulations, but “[w]hat is ‘safe’ for the United States may not be ‘safe enough’ for France or Germany. [...] Nations have different views because they have different preferences and circumstances” (Rodrik, 2011, p. 223). Similarly, foreshadowing a discussion later in this article, while appeals to “universal human rights” may seem to offer the hope of a shared value for the world’s peoples, the legitimacy of particular interpretations will vary greatly across different groups. Tolerance for Holocaust denialism, for example, varies greatly across countries with strong commitments to human rights and free speech (Haggart 2020).

3.4. Interaction effects

These three forms of legitimacy also involve “interaction effects,” in which the quality of inputs, throughputs and outputs taken together can affect perceptions of legitimacy. “Output policy and input participation can involve complementarities and trade-offs, where less quality in the one may be offset by better quality in the other” (Schmidt, 2013, p. 3). For example, one can imagine general acceptance of a human rights-driven content-moderation decision (output) even if it goes against one’s values, if one could argue it was arrived via a process that was seen to be democratically legitimate (input).

When throughput is added to the mix, however, the relationships become more complicated. While good throughput governance has no halo effect, bad throughput governance can delegitimize not only throughput governance, but “can also cast a shadow over both input and output by devaluing even good output policy if it is seen as tainted and more input participation if it is seen as abused” (Schmidt, 2013, p. 9). Governments rarely receive credit for a process that runs smoothly because when they do, they are seen as merely doing what they should be doing. When the process goes awry, voters remember and can judge harshly.

This point is particularly important for our understanding of platform-governance legitimacy, especially given the emphasis placed by so many proposals and activists on transparency (a throughput issue) as a central means to regulating platforms in the public interest.

3.5. A framework for evaluating platform governance

This democratic-legitimacy-evaluative framework, developed to assess the European Union’s democratic bona fides, translates nicely to the field of global platform governance and regulation. It takes seriously the assertion that global platforms represent, or at least aspire to the status of, “the new governors” (Klonick, 2017) and a form of supranational governance. Consequently, we can assess platforms’ democratic legitimacy according to the same criteria. A full understanding of platform-governance legitimacy requires that it be evaluated using all three categories. A platform can score high in one form of legitimacy but low in others.

Input legitimacy with respect to platforms focuses on the ability of citizens in their role as citizens to control the very form of the platform itself, including what the platform does and how it does it. Key here is the question of *accountability* of decision-makers. As such, it is reflected in but not reducible to, the extent to which platforms are responsive to citizens’ demands, in the same way that a monarch responding to the pleas of his subjects would not be confused with a form of democracy. We suggest two criteria:

- The degree to which majoritarian influences are incorporated into platform-governance structures, through representation or participation.
- The presence of a self-identified polity.

Throughput legitimacy refers to the quality of governance processes. Here, we adopt Schmidt's four listed criteria (2013, 6–7) wholesale:

- Transparency – Do people have enough information about decision-making processes and rules to hold rule-makers accountable?
- Efficacy – Does the process actually achieve its intended outcome in an efficient manner?
- Accountability – Are platforms and those working in and for platforms judged (and held accountable; i.e., do they face consequences) for their decisions?
- Inclusiveness and openness to civil-society participation – Are civil-society groups accorded an ongoing role in setting and/or implementing platforms' policies and rules?

Output legitimacy, meanwhile, considers the ultimate result of a platform's activity. It involves two factors:

- the extent to which platform's actions enhance the lives of the people affected by it; and
- the extent to which these actions fit the values and identities of the people affected by said actions.

Finally, the interaction effects of these three forms of legitimacy with each other must be considered. As Schmidt (2013) notes, input and output legitimacy can be complementary: High input legitimacy, for example, may be able to compensate for low output legitimacy. Low input and/or output legitimacy, meanwhile, can negate high throughput legitimacy. Importantly, high throughput legitimacy tends to have a minimal effect on input and/or output legitimacy, while low throughput legitimacy tends to have a negative effect on input and/or output legitimacy.

Understanding these interaction effects is particularly important for evaluating proposed platform-governance policies. Overall, platforms that rely excessively on throughput legitimacy for their overall legitimacy are less preferable than ones that focus on maximizing either input or output legitimacy.

In the following sections, we will apply this framework to four platform-governance policies and proposals.

4. Four paradigmatic case studies

In this section, we apply our framework to four different platform and paradigmatic governance proposals: Facebook's Oversight Board (self-governance regimes); adjudication-focused proposals such as the *Manila Principles for Intermediary Liability* (rule-of-law-focused regimes); the human-rights-focused framework proposed by the then-UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (international law regime); and the United Kingdom's Online Harms White Paper (domestic regimes).

4.1. Facebook's Oversight Board

In November 2020, Facebook's Oversight Board, which had been under development since November 2018, began operations. According to Facebook, its purpose is to serve as an independent appeal body with competency to review Facebook's content moderation decisions, with the goal of protecting "free expression by making principled, independent decisions about important pieces of content and by issuing policy advisory opinions on Facebook's content policies" (Facebook, 2020, p. 5). Ultimately, it purports to provide a solution to Facebook's current "unchecked system" for users' speech governance (Klonick, 2020, p. 2476; Douek, 2020, p. 46).

It is too early to make a conclusive determination as to whether the Oversight Board will exert either meaningful independence or prove to be a check on Facebook's actions - its first decisions were announced in January 2021 (Oversight Board, 2021). Consequently, we focus on its purposes and processes as set out in Facebook's official documents and academic research into its design. This information, however, is sufficient to make some preliminary comments about the Board's overall legitimacy.

Originally, the Oversight Board would only analyze appeals regarding content that was removed based on Facebook's own rules. Two months after its first decisions, the Oversight Board announced it would be expanding its scope to also accept claims regarding Facebook's decisions to keep content online; i.e., publications that users feel should be removed (Oversight Board, 2021a). Either way, it will deliberate on content disputes according to the website's Community Standards. This excludes from its appreciation all content taken down under claims of incompatibility with local laws. This restriction on the Board's scope can be understood as a natural recognition that such a body could not claim functions of legal interpretation and of its limited power to mandate non-compliance with legal requirements (Douek, 2019, p. 38) - even though Facebook still acts as a private enforcer when it decides compatibility with the law in the first place.

Regardless of its merit on principle, the decision regarding scope is fundamental in order to properly evaluate the importance and potential influence of this Oversight Board in the greater online content governance landscape. Within this already-restricted scope, the possibility for appeals regarding content taken down according to the Community Standards is minimal given the "sheer volume of content moderation decisions that Facebook makes every day": the the Board "cannot be expected to offer this kind of procedural

recourse or error correction in anything but the smallest fraction of these cases” (Douek, 2019, p. 3).

4.1.1. Input legitimacy

The claim for input legitimacy on the part of the Oversight Board is a weak one, given that it is a body instituted by a private company and implemented with no democratic oversight. Because it is conceptualized, operationalized, and funded by a private corporation, it clearly stands in the realm of corporate *self-regulation*.

From Klonick’s description of the Board’s conception, one can perceive on behalf of Facebook’s the intention to promote extensive participation in the process of the Board’s design, possibly as an attempt to improve its input legitimacy (2020, p. 2451). Klonick describes a global consultation process that “would reach out to stakeholders and users worldwide to survey what people thought the Oversight Board should look like, accomplish, and address” (Klonick, 2020, p. 2451). As a result, “650 people from eighty-eight countries attended the workshops ... [and] 1,206 people participated in the online questionnaire (...)” (Klonick, 2020, p. 2456). While these are large numbers, in the face of the 2.5 billion Facebook users, this rate of participation does not exactly meet a sufficiently representative standard to secure input legitimacy.

Further, neither this consultation process nor the structure of the Oversight Board addressed the lack of user participation in the creation and updating of the Community Standards, which precede the Board, and had already drawn the boundaries for the content moderation landscape (Klonick, 2020, p. 2424). Regarding the question of “who makes the rules?”, not much has changed: Facebook is still clearly in charge.

4.1.2. Throughput legitimacy

Throughput legitimacy offers the strongest potential legitimacy claim of the Oversight Board, because it focuses on inserting transparency, due process, and accountability into content moderation decisions.

As put by Klonick, the Oversight Board has a “robust level of transparency” (2020, p. 2480), including: publication of the applicable rules; notification of infringement and review procedure; explanation of what this process entails; and notification of the ultimate decision (Klonick, 2020, pp. 2479–2480). Also, its Bylaws commit the Board to making all case decisions publicly available, archiving them in a database and publishing annual reports with metrics on the cases reviewed, cases submissions by region, timelines of decisions. The Board’s potential for accountability, on the other hand, is less promising, given that its decisions are not subject to further scrutiny.

Nevertheless, it can make a strong claim for legitimacy with respect to due process. Due process is in fact perceived as one of the Oversight Board’s defining characteristics (Douek, 2019, p. 6). The central goal of the Board is to grant Facebook users the possibility of having their content controversies examined by a selection of experts from different world regions who are allegedly independent from Facebook.

As much as throughput legitimacy might be the Oversight Board’s strongest claim to democratic legitimacy, its standards are not completely fulfilled. Openness to interests’ intermediation, for instance, is restricted to the possibility of appeal. There are no provisions for other sorts of formal participation mechanisms.

Lastly, assessing and evaluating the Board’s efficacy requires consideration of what one understands as “functioning”. The Board’s Bylaws do not offer much insight on this question. Its Introduction consisely states its goals: “[t]he purpose of the Oversight Board is to protect freedom of expression by making principled, independent decisions about important pieces of content and by issuing policy advisory opinions on Facebook’s content policies”. These seem like laudable goals. However, as we have already noted, the sheer amount of content produced by Facebook users, as well as its geographic scope, raise significant and potentially insurmountable obstacles to its ability to efficiently achieve its objectives.

Instituting an effective and throughput-legitimate framework—especially when its scope is so restricted—will depend on how this operation is implemented; if the process indeed involves thorough independent reviews, providing a “way of ensuring the effective functioning of a bureaucratic system and rule enforcement by creating a mechanism for error correction “given Facebook’s enormous size and scope” (Douek, 2019, p. 6).

4.1.3. Output legitimacy

The Oversight Board possesses questionable output legitimacy, mostly due to its limited scope, long timelines for decision making, and the fact that the company’s set of values and principles (which informs the Board) will inevitably clash with those of various national regimes and affected groups.

Besides its already described restricted scope, Board decisions risk being undermined by potentially lengthy deadlines. The Board only receives cases after users have exhausted the internal appeal processes; from the moment that Facebook makes its first take-down decision, the Board has 90 days to analyze and reach theirs (Klonick, 2020, p. 2471).

Moreover, Facebook’s rules and values which provide the foundation for the Board’s operation are not neutral (since they are embedded within the company’s own interpretation of the scope of free speech), nor can the Board “draw on universal categories because no such consensus exists” (Klonick, 2020, p. 2475). Even though the Board’s decisions provide users with an extra, expertise-driven level of scrutiny, they are still likely to clash with others’ norms and expectations (Klonick, 2020, p. 2475).

4.1.4. Interactions and overall assessment

The Oversight Board’s design emphasizes throughput legitimacy. It was conceived as a means to implement adjudication performed by experts, while also ensuring meaningful transparency regarding the reasoning process underlying specific content moderation decisions (Klonick, 2019). At the end of the day, the Board is an instrumental mechanism designed to ensure procedural

guarantees for a small fraction of online content disputes. It does not have input legitimacy, while its output legitimacy is likely to be questionable. Its biggest claim to legitimacy therefore is throughput legitimacy, meaning that it provides an "independent" instance to rule on content-removal decisions, and that it is committed to abiding by pre-established procedures.

The Board does not propose a new normative framework that could exercise a significant influence on Facebook's output; although it can suggest changes to Facebook, the Community Standards and Values that ideological bind its decisions are still the same ones that have guided Facebook's content moderation. It also fails to facilitate or enhance control of content moderation by democratic institutions, nor does it even implement significant participation instruments in Facebook's decision-making and norm-setting processes (regardless of the global consultation process).

In terms of overall legitimacy, the main concern with the Oversight Board is that robust throughput may not be enough to compensate for the lack of input and output legitimacy. Our assessment strongly suggests that the Board is likely to be a procedure-focused institutional arrangement that has the potential to implement a more constitutionally responsive online content moderation process but remains considerably limited by its scope. This design is certainly defensible, but it is also very narrowly focused. However grand Facebook's intentions for the Oversight Board, we should not expect it to address Facebook's overall problems of legitimacy surrounding content moderation.

4.2. Judicial adjudication

Overall, court adjudication provides legal certainty to the interpretation of the law in concrete cases. For platform governance, it plays an important role in defining the scope of freedom of expression in the face of harmful conduct. However, the over-reliance on courts as a primary platform-governance mechanism may raise certain issues, discussed below.

Adjudication is one of the *Manila Principles for Intermediary Liability*, which were "developed by an open, collaborative process conducted by a broad coalition of civil society groups and experts from around the world" ([Manila Principles - Background Document 2015](#)), aiming to develop "intermediary liability policies that can foster and protect a free and open Internet" ([Manila Principles, Background Document, 2015](#)). Court adjudication is its second principle, according to which "[c]ontent must not be required to be restricted without an order by a judicial authority" ([Manila Principles, 2015](#)). It recommends the adoption of systems where any liability imposed on intermediaries is directly correlated to their wrongful behavior in failing to comply with the content restriction order ([Manila Principles, 2015](#)).

It has also been identified as a legitimacy standard of intermediary liability in UN Specialist reports ([de la Rue 2013](#); [Kaye, 2018](#)) and in the academic literature (Urs and Schulz, 2015). Some countries have embraced the Principles' call for mandatory, court-overseen liability limitation as a democratic practice (Gasser and Schulz, 2015) for its alleged privileging of freedom of expression ([Keller, 2020](#)). The assumption underlying the embrace of a strong role for judicial authorities is that by creating incentives for liability claims to be brought to the judiciary, platforms will be less likely to remove content simply because a notification has been received (Lemos and Souza, 2015), thus reducing platforms' incentives to overblock ([Kaye, 2018](#)). Conversely, a full liability regime (e. g., where financial reparations by the platform are due upon mere extrajudicial notification) would encourage private monitoring and exclusion of potentially controversial material, threatening the circulation of legitimate content.

Beyond the Manila Principles, this type of regime has been adopted by Brazil in its *Marco Civil da Internet* internet-governance legal framework. Under section 19 of the *Marco Civil*, intermediaries are only liable for third-party harmful content if they fail to remove content after receiving a judicial take-down notification. Instead of being potentially responsible for any committed (or acknowledged) infringement, findings of liability rest on the assessment and interpretation of judges. A similar system has been adopted by Chile, where actual knowledge of infringing content (required for liability) is equated to judicial notification ([Lara & Vera, 2013](#)).

The assumption that court-centered intermediary liability privileges freedom of expression has already been challenged for its systemic effects on online content governance, referable to the judiciary's institutional design ([Keller, 2020](#)). Yet, handing these decisions to courts would, in theory, address the legitimacy deficit faced by most decision-making processes in the realm of platform governance, since courts rely, implicitly and to a certain degree, on democratic legitimacy to legitimize their own activities.

4.2.1. Input legitimacy

At first glance, input legitimacy might be considered the strongest claim for submitting online content disputes to formal judicial adjudication. Even though courts are non-elected bodies, they are the product of institutional designs rooted in political representation, operating "in the 'shadow of politics', as the product of political institutions" ([Schmidt, 2013](#)). Control over "the precise meaning of the test for restrictions on freedom of expression" (Mendel, 2010) is at the heart of their *raison d'être*. However, their democratic legitimacy only holds to the extent that the state in which they operate is also democratic.

Nevertheless, this does not mean that there are no issues regarding the input legitimacy of judicial adjudication processes. Adjudication of intermediary liability can generate "policy-like" effects on a country's online content-governance regime, bypassing the governmental branches that are nominally responsible for implementing public policies ([Keller, 2020](#), pp. 16–17). Such situations, where the judicial branch effectively embraces legislative-like powers, could result in a loss of the court's input legitimacy, when they move from adjudicating to policy-making ([Keller, 2020](#)). Such a result can be problematic considering that the constitutional law literature has already recognized that the judicial branch is not necessarily reflective of the flow of popular representation that sustains public policy choices (Hirschl, 2008).

Because freedom of expression is both an individual right and a collective guarantee ([Machado, 2001](#)), its promotion also depends on positive state provisions that are inherently under the legislative and original executive competences. As a result, when policies focus too much on the courts at the expense of the legislative and the executive, the shadow of politics is lifted and their legitimacy can

be brought into question.

4.2.2. Throughput legitimacy

Initially, adjudication would also seem to possess strong throughput legitimacy. Regarding transparency, lawsuits are usually presumed to be public, except for those cases that are examined under secrecy. Due process is a given (at least in democratic states) and accountability is assured by the possibility of appeal, besides other institutions with competency to investigate complaints or misconduct.

That said, the qualities of openness and efficacy in adjudication processes remain highly influenced by the regulations governing how lawsuits are conducted. An essential presumption underpinning these two factors is that judges will be able to access all the relevant information necessary in order to reach optimal decisions. Nevertheless, “relevant information” encompasses only the information regarding the specific dispute to be adjudicated. Even though the legal pragmatism school advocates that judges take possibly concrete consequences of their decisions into consideration (Pereira, 2016), their reasoning process is bound by the facts, arguments, and legal sources that are brought by the parties to the records.

Accordingly, these limitations would make the judiciary unfit for adjudicating policy—as is the case for platform governance. They concern the “efficacy” dimension of throughput legitimacy, since adjudication has consequences for the greater online content governance framework that is affected by single-case-based decisions. Efficacy is also likely to be affected by the characteristically slow speed of judicial processes, challenged to provide solutions to a technological dynamic context in accordance to its own bureaucratically paced rhythm.

Openness can similarly be influenced by the way lawsuits are conducted, as all information to be considered is what is presented by the involved parties. It should be noticed that this can be mitigated in systems that allow for *amicus curiae* (friend of the court) briefs in certain lawsuits. Yet openness is also affected by *selectivity*, as courts act upon individual demands that are brought to them by single claims. This means that a series of cases that were not brought to courts (by choice or from a lack of material means) are left outside this scrutiny mechanism. In fact, this affects openness (as it affects access) and output, as described below.

4.2.3. Output legitimacy

More positively, the output legitimacy of adjudication-focused policies are aided by the fact that court decisions are binding and provide a high degree of certainty in resolving disputes. The assumption here is that court decisions represent a decisive interpretation of how law applies to that case and represent an interpretation by which all parties must abide.

Conversely, however, the fact that judges often lack the expertise to adjudicate matters that require specific technical knowledge, plus the lawsuit scope limitations mentioned above, can favor decisions that disproportionately restrict rights, are simply ineffective, or both. Such decisions were frequent in the Brazilian system, leading to a situation where single-case adjudication can generate collectively harmful decisions. To take only one example, in the lawsuit known as *Cicarelli case*, a claim made by a famous TV personality and her boyfriend against YouTube resulted in a 48-hour-long imposed blockages on this website (Porfirio, 2016). The couple had sued the platform demanding the removal of an unauthorized video, as well as damages for the violation of privacy. After an initial decision mandating YouTube to remove the video proved “trickier than expected” as “[co]pies of the original video kept resurfacing on the platform” (Moncau e Arguelhes, 2020), the trial judge ordered several internet access providers to block the YouTube platform itself on the whole Brazilian territory. Decisions such as these are clearly oblivious both to the fact that they can immediately affect millions of users’ access-to-information rights while ignoring the fact that the exact same content can immediately be available in numerous other sources.¹ An individual case becomes a *de facto* policy decision based on a failure to “anticipate the secondary effects of their judgments” (Gasser and Schulz, 2015).

Lastly, as mentioned above, selectivity in cases heard can also affect output legitimacy: courts lack the capacity to deal with every platform-governance issue. One of the consequences of this lack of capacity is that embracing these principles would still result in a significant proportion of online content regulation effectively being left to the discretion of intermediaries who would then be free to apply their own product’s policies in matters of content moderation, according to their own criteria.

4.2.4. Interactions and overall assessment

The throughput and overall legitimacy of adjudication-focused approaches in online content disputes is largely dependent on how the tensions between the individual and collective dimensions in which these cases are placed are navigated. Individually, each case represents one complaint requiring content removal that will be appreciated by a judge, who will rule on the scope of freedom of expression. Judges’ decisions will be presumed to be legitimate, enforceable and legally certain—even if their content is oblivious to technical expertise. In this sense, high levels of input legitimacy compensate for (possibly) poor degrees of output legitimacy.

However, as suggested by the above analysis of output legitimacy, these decisions will inevitably shape the platform policy-making context, possibly generating systemic effects that are beyond the judicial process’s capabilities to anticipate or manage. When one considers Schmidt’s assertion that deficiencies in throughput legitimacy will tend to have a negative effect upon output legitimacy, these deficiencies suggest the dangers of relying on the judiciary, even of a democratic country, to play too-great a role in platform governance.

Nonetheless, adjudication, particularly in a democracy, can play an important role in protecting individual rights; similarly, the

¹ For a detailed description of this case and others, see Moncau and Arguelhes (2020).

intermediary liability standard discussed in this section is a defensible policy choice *on its own*. However, we emphasize that, on its own, judicial adjudication is not enough to address the complexity of platform governance dynamics. In addition to the throughput limitations that narrow lawsuits, they are also slow and inherently selective; only those people with the financial and subjective means to sue these companies will get public interest scrutiny over their online content dispute.

4.3. A human-rights-centric framework

In a 2018 United Nations report, David Kaye, UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, “proposes a framework for the moderation of user-generated online content that puts human rights at the very centre” (Kaye, 2018, paragraph 2). Although Kaye “acknowledge[s] the interdependence of rights, such as the importance of privacy as a gateway to freedom of expression” his proposal focuses exclusively on the right to “freedom of expression” (Kaye, 2018, paragraph 5).

In Kaye’s account, the current treatment of online speech is negatively affected by both state and platform actions. Platforms suffer from vague, profit-motivated, inconsistent and unclear rule-setting, all too often shaped by parochial home-country (usually US) biases: “Private norms, which vary according to each company’s business model and vague assertions of community interests, have created unstable, unpredictable and unsafe environments for users and intensified government scrutiny” (Kaye, 2018, paragraph 41). States, meanwhile, are problematic because of vague laws “subject to varying interpretations or inconsistent with human rights law” or when they put pressure on companies “to comply with State laws that criminalize content that is said to be, for instance, blasphemous, critical of the State, defamatory of public officials or false” (Kaye, 2018, paragraph 23).

Kaye’s human rights-centric framework emphasizes the implementation of “human rights standards (...) implemented transparently and consistently, with meaningful user and civil society input” as a way to hold “both States and companies accountable to users across national borders” (Kaye, 2018, paragraph 41). In particular, the report emphasizes the following:

- The adoption by companies of “high-level policy commitments to maintain platforms for users to develop opinions, express themselves freely and access information of all kinds in a manner consistent with human rights law.”
- Transparency in company decision-making.
- User and civil society input to, among other things, “consider the human rights impact of their activities from diverse perspectives,” and to audit algorithmic technologies.

4.3.1. Input legitimacy

The extent of an institution’s democratic input legitimacy, as we have noted, reflects its “responsiveness to citizen concerns as a result of participation by the people” (Schmidt, 2013, p. 5). While advocating for international human rights as the framework within which platform policies can be discussed and critiqued, this proposal does not spend much time on the democratic question, namely, to whom should a platform be ultimately accountable. That is, who should govern?

In a section titled “Rule-making and product development,” Kaye calls on companies to “develop clear and specific criteria for identifying activities” that put freedom of expression at risk (Kaye, 2019, paragraph 53). Development of these policies, Kaye argues, should be done with the assistance of “multi-stakeholder initiatives such as Global Network Initiative,” as well as “affected rights holders (or their representatives) and relevant local or subject matter experts,” and with the involvement of “users and civil society, particularly in the global South” (Kaye, 2018, paragraphs 53 and 54).

While such an approach may seem to possess at least some democratic input legitimacy, it faces two significant problems, both related to the wide range of interpretation possible when it comes to operationalizing universal human rights, or international law generally. First is the long-recognized question of the representativeness of civil society organizations, including with respect to internet multistakeholder governance (Carr, 2015; Powers & Jablonski, 2015). Such groups, no matter how well-intentioned, do not draw their authority from any democratic mandate. Furthermore, as Kaye implicitly recognizes, human rights standards are highly contestable (Kaye, 2018, paragraph 41). Human rights law’s openness to different good-faith interpretation means that such groups also cannot claim legitimacy through technical expertise (Rodrik, 2011, p. 223). Kaye does not consider the fundamental question of civil society’s or the user community’s representativeness. Even the smallest society contains divergent views on speech issues. Given this reality, to whom should we listen? Kaye provides us with no guidelines for how to determine how different groups’ interests should be prioritized.

Second, although Kaye clearly recognizes the problems that arise when companies are left to set their own rules, his framework fails to challenge their supremacy. Kaye’s approach to the question of who should set the rules is nicely captured by his comment that “when companies develop or modify policies or products, they should actively seek and take into account the concerns of communities historically at risk of censorship and discrimination” (Kaye, 2018, paragraph 48). This approach is closer to the monarch’s *noblesse oblige* than it is to a “new governor” that is democratically responsive to the desires of a citizenry. It is also hard to see how it differs much from the *status quo*: As Jorgensen (2017, p. 280) notes, Facebook and Google both “frame themselves as strongly committed to (and actively promoting) human rights,” although they tend to focus on government violations while paying less attention to their own actions and “actually retain the freedom to set and enforce their own rules of engagement.”

Of course, there exists a process and actor for adjudicating questions such as where to draw the line between acceptable and unacceptable speech: the democratic state via elections. However, Kaye’s framework views all states with suspicion; he calls on companies to “develop tools that prevent or mitigate the human rights risks caused by national laws or demands inconsistent with

international standards” (Kaye, 2018, paragraph 49). Again, the relevant question is, who ultimately decides when a law is inconsistent with international standards? Whether intentionally or not, Kaye’s answer is, effectively, the platform.

By treating all states effectively as illegitimate rule-setters, Kaye also removes from consideration the one existing mechanism for providing platform regulation with any kind of democratic legitimacy.

4.3.2. *Throughput legitimacy*

In the absence of input legitimacy, Kaye’s framework is rooted primarily in throughput legitimacy and a flawed approach to output legitimacy. Turning first to throughput legitimacy, Kaye focuses on issues of transparency, inclusiveness and openness to intermediation; questions of efficacy are not highlighted in his proposal. In particular, with “due diligence, transparency, accountability, and remediation that limit platform interference with human rights” (Kaye, 2018, paragraph 42) central to the functioning of his proposed system. Throughput legitimacy can also be seen in Kaye’s call for “the professionalization of their human evaluation of flagged content, remediation processes, and the provision of user autonomy (Kaye, 2018, paragraphs 57, 59, 60). Transparency in rule-making and in justifying decisions, is treated almost as the cornerstone of this framework, enabling users to assess the company’s actions (Kaye, 2018, paragraphs 55, 61, 62, 63).

Throughput legitimacy also involves consultation. It could be argued that Kaye’s attempts to involve users and civil society (e.g., in paragraphs 53–55) should be more appropriately be seen as a way of improving throughput processes, rather than increasing input legitimacy. The framework would invite civil society to provide input into algorithmic reviews and policy development, including in ways that encourage companies “to pay close attention to how seemingly benign or ostensibly ‘community-friendly’ rules may have significant, ‘hyper-local’ impacts on communities and algorithmic review” (Kaye, 2018, paragraph 54). The key word, again, is consultation: companies remain in charge of the process.

4.3.3. *Output legitimacy*

Kaye’s framework attempts to draw its legitimacy in terms of its desired output, the consistent, universal application of international human rights law. Unfortunately, while international human rights law can provide a useful framework for a conversation about proper regulation, it is insufficient on its own to provide provision of output legitimacy. As noted above, these high-level standards must be operationalized via specific laws, rules, regulations, terms of service and norms. This operationalization is at least as important as the choice to ground specific regulations in human-rights law. As Kaye himself notes, Article 19 (3) of the UDHR allows for restrictions based on “the rights or reputations of others, national security or public order, or public health or morals” (Kaye, 2018, paragraph 7). These exceptions – which are anything but trivial – hint at the reality that free speech as a norm is necessarily in tension with other rights.

As Schmidt (2013) notes in the context of the EU, policies that fail to take into account local differences will lack output legitimacy. Kaye’s framework fails to provide sufficient attention to the complexity of local differences. A statement like the following – “Companies should ... adopt high-level policy commitments to maintain platforms for users to develop opinions, express themselves freely and access information of all kinds in a manner consistent with human rights law” (Kaye, 2018, paragraph 45) – applied to the real world assumes a universal hierarchy of values and an unproblematic view of the application of (international) “human rights law” than is warranted in a pluralist world.

That reasonable people can disagree about where to draw lines regarding speech restrictions highlights the fundamental legitimacy flaw in Kaye’s proposal: it assumes that the regulatory question is, “Where should we draw the line?” In the absence of objective criteria, the real problem is, “Who should be responsible for drawing the line, and how should they decide?”

4.3.4. *Interactions and overall assessment*

This assessment suggests that Kaye’s human rights-focused framework does not draw its primary legitimacy from its promise of delivering respect for human rights (output legitimacy). With respect to democratic input legitimacy, its sidelining of democratic states, failure to assess the democratic bona fides of “users” and “civil society” as public representatives, and failure to challenge companies’ ultimate rule-setting responsibility, mean that it also lacks strong democratic input legitimacy. Again, international human rights law may be a useful frame for discussing platform governance, but its operationalization remains the central issue. Beyond this appeal to human rights norms, Kaye’s framework focuses to a significant degree on throughput legitimacy particularly as it relates to transparency. (Actual accountability, meanwhile, is underdeveloped in the paper.) As Schmidt (2013) suggests, this is a very thin reed upon which to base a governance regime: Lacking meaningful input or buy-in on the output, the fact that a platform is more open about how it came to make unpopular decisions is unlikely to win it any support among the people over which it rules.

4.4. *United Kingdom Online Harms White Paper*

The United Kingdom released its Online Harms White Paper in April 2019 (Department for Digital, Culture, Media & Sport and Home Office, 2019; hereafter White Paper, 2020). It was followed up by a government response in December 2020 (Department for Digital, Culture, Media & Sport and Home Office, 2020, 2020a; hereafter Government Response 2020) and legislation is expected sometime in 2021, although it may not come into force until 2022 or later (Kelion, 2020). The White Paper and governmental response propose a domestic regulatory framework to address online “illegal and unacceptable content and activity” (White Paper, 2019, p. 5), to be overseen by the UK’s independent telecoms regulator, Ofcom (Government Response, 2020, p. 5).

Of particular interest, the White Paper proposes the establishment of a statutory “duty of care” (Government Response, 2020, p. 9) for online companies similar to the “high-level policy commitments” proposed by Kaye. Ofcom, the UK communications regulator,

would be responsible for setting out “codes of practice” detailing how companies should fulfill this new duty (Government Response, 2020, p. 23), while also being expected to put “systems and processes in place” to “improve user safety on their services,” including reporting and redress mechanisms (Government Response, 2020, p. 27).

Ofcom, meanwhile, would also be responsible for addressing both “illegal” and “harmful” content that is not necessarily illegal (see Table 1) – “harmful content and activity” to be defined generally in legislation, but generally applying to that which “gives rise to a reasonably foreseeable risk of a significant adverse physical or psychological impact on individuals” (Government Response, 2020, p. 9; p. 24). The definition of “illegal content” is determined by statute, the definition of “legal but harmful content” is to be “set out in secondary legislation” (Government Response, 2020, p. 23). Table 1 below outlines the White Paper’s proposed distinctions among legal and illegal harmful content; the 2020 government response commits to a legislative definition of harmful content and activity, rather than leaving it to the regulator to interpret (Government Response, 2020, p. 24).

Source: Online Harms White Paper, 2019, p. 31.

Companies under this framework would be required to publish “annual transparency reports,” with “information about the steps that they are taking to tackle online harms” on their services, with Ofcom given the power to determine the content of these reports (Government Response, 2020, p. 67–68). Companies would also be required to provide complaints mechanism; Ofcom would also respond to “super complaints” related to systemic issues, and “establish ongoing mechanisms for user advocacy” (Government Response, 2020, p. 71–72). Ofcom’s decisions would be subject to appeal (Government Response, 2020, p. 76). Ofcom would also be required to produce a report on the extent to which independent researchers are being given access to company data “to support research into online harms” (Government Response, 2020, p. 70). These transparency requirements are designed to increase platform accountability to regulators and citizens. Companies would also be required to have “effective and easy-to-access user complaints functions, which will be overseen by the regulator,” with the additional possibility of a designated overarching “super complaints” body (White Paper, 2019, p. 8).

4.4.1. Input legitimacy

As a domestic policy response, the White Paper is firmly embedded within a well-established democratic policy making process. The initial draft White Paper outlined the Conservative government’s proposal. It was then subject to “a formal 12-week consultation” between April and July 2019. Some “2,439 respondents from across academia, civil society, industry and the general public” addressed 18 questions (12 of which were open-ended) that comprised the written phase of the consultations; the government also engaged in 100 meetings with stakeholders (Government Response, 2020, pp. 5–6; White Paper, 2019, pp. 97–98). With respect to ongoing rule-making, the “duty of care,” as well as definitions of illegal and harmful activity will be set out in parliamentary statute.

4.4.2. Throughput legitimacy

While it is impossible to judge some aspects of throughput legitimacy – particularly efficacy – at this stage, we can offer an initial assessment of the extent to which the White Paper’s regulatory model addresses some of these issues. The White Paper and governmental response pays significant attention to transparency as a way to improve regulatory quality and platform-user understanding – the governmental response dedicates an entire section to it. It also claims that the regulator will adopt “an evidence-based approach to regulatory activity” (2019, p. 56) drawing on expert advice, including “users, civil society, government, law enforcement and other relevant government agencies, and other regulators to inform its understanding of the prevalence and impact of online harms, and the effectiveness of companies’ responses” (2019, p. 45). With respect to ensuring its work is based on research, the government expects “the regulator will continue to support research in this area” (White Paper, 2019, p. 27). These moves, reinforced by the Government Response’s mentions of research as an “essential part” of Ofcom’s role (2020, p. 71), suggest at least a nod toward inclusiveness in the policy-setting process.

Like other independent regulators this independent regulator would not be directly accountable to the citizenry, but rather bound

Table 1
Online harms in scope.

Harms with a clear definition	Harms with a less clear definition	Underage exposure to legal content
Child sexual exploitation and abuse.	Cyberbullying and trolling. Children accessing pornography.	
Terrorist content and activity.	Extremist content and activity.	Children accessing inappropriate material (including under 13s using social media and under 18s using dating apps; excessive screen time
Organized immigration crime.	Coercive behavior.	Extremist content and activity.
Modern slavery.	Intimidation.	
Extreme pornography.	Disinformation.	
Revenge pornography.	Violent content.	
Harassment and cyberstalking.	Advocacy of self-harm.	
Hate crime.	Promotion of Female Genital Mutilation (FGM).	
Encouraging or assisting suicide.		
Incitement of violence.		
Sale of illegal goods/services, such as drugs and weapons (on the open internet).		
Content illegally uploaded from prisons.		

by participation mechanisms (i.e., public consultations) and other legitimacy enhancing features, besides the fact that it would presumably fall under a Cabinet ministry. That said, the creation of a user-redress system should introduce a degree of accountability into the platform-governance process. However, as the regulator has yet to be created, these remarks are necessarily preliminary.

4.4.3. Output legitimacy

As the White Paper has yet to be implemented, its output legitimacy – the extent to which the eventual process's outcomes are taken as legitimate by UK citizens – is necessarily unknowable. The government's *Initial Consultation Response*, however, provides some hints as to what we can expect in this area. It found that a "notable number of individual respondents ... reiterated a general disagreement with the overall approach" (2020a). These individuals are highly unlikely to see as legitimate any outcome from whatever program the government proposes.

For the remainder of the population, output legitimacy will likely depend on their assessment of the extent to which they feel that the eventual government response is able to deliver outcomes in line with their values and opinions. Possible influences on their reaction could include the extent to which illegal or perceived harmful online behaviours are actually reduced.

4.4.4. Interactions and overall assessment

The White Paper, based in a domestic, longstanding and widely accepted democratic framework, can make strong claims to input legitimacy. Similarly, while its actual effectiveness – in terms of process and outcomes – remain to be tested, its design addresses throughput legitimacy issues such as accountability, transparency, inclusiveness and openness.

With respect to input legitimacy, the independent regulator would benefit from being "in the shadow of politics," through its creation by a democratically legitimate government. The democratic input legitimacy of the proposal, meanwhile, is further strengthened by its treatment of legal but harmful content and activity. While the White Paper initially left the definition of this category to the regulator, basing it more firmly in legislation, as the governmental response proposes, further increases its democratic bona fides.

5. Four platform-governance legitimacy lessons

As the preceding four case studies show, decomposing the concept of policy legitimacy into its component parts offers us several useful insights into how the concept of legitimacy is deployed in platform-governance proposals. We highlight four in particular.

5.1. Emphasis on throughput legitimacy (with one exception)

Non-state-focused proposals (i.e., all but the UK White Paper) are based on a narrow conception of legitimacy as throughput legitimacy. In particular, all three emphasize the importance of transparency and process-oriented characteristics such as due process guarantees.

While process is certainly important, that legitimacy has been so narrowly interpreted should be concerning for two reasons. First, as we note above, in some cases, the focus on process may not be sufficient even on its own terms. With respect to adjudication-focused processes, for example, that court procedures are presumed to be accountable, transparent and follow due process would seem at first to provide a strong claim to significant throughput legitimacy. Nevertheless, the institutional design of lawsuits gives them limited capacities to assess, and therefore, adjudicate on, big-picture policy issues such as internet governance. Second, and more importantly, as Schmidt remarks, throughput legitimacy is a weak foundation upon which to build a legitimate institution: robust processes contribute little to overall legitimacy (because it's taken as given that they should run well), while flawed processes can discredit an entire project. As a result, we conclude that, even though throughput concerns are important, proposals and policies that base their legitimacy claims primarily on the quality of their throughput are likely to face difficulties in credibly asserting overall legitimacy.

5.2. Input legitimacy: who makes the decisions question is underexamined

This overemphasis on throughput legitimacy in non-state-focused proposals and policies is paired with an equivalent lack of consideration of who makes the decisions. Given that Facebook's Oversight Board will only adjudicate community standards (and in limited situations), its normative framework remains Facebook's internally and non-transparently set rules. The Board's structure does not promote any change or enhance participation of any sort in its private rule-making processes. As a throughput-focused entity, it essentially plays the procedural role of an appeal body. Adjudication processes, meanwhile, do not necessarily have suitable institutional characteristics to decide policy issues. Lawsuits are presumed to only contain the information brought by the parties involved, referable to the rights conflict of each case. Instead, they, like the proposed regulator in the UK Online Harms White Paper, draw their input legitimacy from the "shadow of politics." However, such courts have significant institutional limitations in what comes to adjudication of public policy fields, which have already been flagged by contemporary constitutional law literature in different regulatory fields. These institutional limitations are intrinsically related to the assessment of input legitimacy, and they have great potential to negatively affect the adjudication of complex policy fields – which is exactly the case for platform governance. However, these constraints are rarely, if ever, mentioned in discussions of adjudication-focused proposals. Finally, with respect to Kaye's proposal, the state appears more as the actor that needs to be restrained than as a positive regulator, while civil society is treated as an advisor. The end result, as noted above, the platform – which lacks democratic input legitimacy – remains the dominant rule-maker (Haggart, 2020).

In contrast, the UK Online Harms White Paper, situated within the context of a longstanding democratic process, can claim strong democratic legitimacy in its development. Furthermore, depending on how it is actually rolled out, the delegation of Parliamentary powers to a quasi-independent regulatory body similarly can claim democratic legitimacy through the “shadow of politics”.

5.3. *The sidelining of state regulation as a policy option*

This decomposing of legitimacy into its component parts also highlights the extent to which the non-state-based proposals and policies all downplay the role of the state as a potential regulator. That Facebook’s Oversight Board is not an instrument of the state is obvious. More interestingly, however, is the relative neglect of the role of the state in the Kaye and adjudication cases (more specifically here, the state’s legislative and executive functions). For example, an adjudication-focused approach to platform regulation, as noted above, focuses on process issues. While these are certainly important to a well-functioning governance regime, they on their own cannot deliver legitimacy. Furthermore, as Keller (2020) notes, courts alone are not competent to design and implement policy, as they lack the expertise and institutional capacity to address complex regulatory contexts, whose scope goes beyond the concrete cases that are presented before courts.

Meanwhile, Kaye’s proposal includes a standard of responsibility based in international human rights law that is practically equivalent to the UK’s proposed “duty of care.” However, Kaye’s deployment of international human rights law is largely aimed at restraining overly zealous state free-speech restrictions. Crucially, neither this proposal nor its book-length elaboration (Kaye, 2019) differentiate significantly between democratic and non-democratic states when it comes to the legitimacy of their regulatory efforts. From a democratic input-legitimacy perspective, one cannot place on the same level the regulatory efforts of a democratic country like New Zealand and those of an authoritarian dictatorship like China. Doing so effectively undermines the notion of democratic-state platform regulation.

5.4. *Reconsidering the role of the state*

Because platform regulation is a relatively new area, and given that we are primarily discussing proposals, there is little concrete we can say about the legitimacy of the outputs of these proposals and policies. And in the case of throughput-focused adjudication processes, there is little to be said about output. That said, there is reason to be concerned about whether a globally focused framework can achieve the buy-in from so many diverse communities needed to achieve output legitimacy. This is not just the case for Facebook’s Oversight Board, but also for Kaye’s human-rights-focused approach. More important than basing a policy framework in international human rights law is how these laws will be interpreted and operationalized; legitimacy in this case will require grounding in accepted interpretations of human rights norms, and these will vary by region and social group.

Overall, this analysis of paradigmatic cases highlights the extent to which the state-based option – the UK Online Harms White Paper – stands out as the one most deeply grounded in input, throughput and output (at least prospectively) legitimacy. This is not to say that it is a perfect proposal, and its implementation may leave much to be desired. However, if one takes seriously the democratic norm that individuals have the right to set and influence the rules under which they must live, one cannot easily dismiss it, particularly in light of the legitimacy flaws of other non-state-based proposals.

This overall analysis highlights the reality that the democratic state remains the entity most able to deliver accountability and legitimacy to its citizens. Without democratic accountability, constitutional forms at the transnational level lack legitimacy. This is not to say that transnational democratic forms may not arise in the future, nor is it to say that the state is the only “level” at which legitimate platform regulation can occur. For now, however, the importance of the democratic state – both on its own and working multilaterally with like-minded democratic states – should not be understated.

6. Conclusion

The question of what constitutes legitimate frameworks for regulating online platforms will doubtlessly remain a central topic in platform-, internet- and global-governance policy circles for the foreseeable future. In this article, we have demonstrated how focusing on democratic legitimacy – itself an under-considered issue in platform-governance circles – as a multifaceted phenomenon can highlight underappreciated weaknesses and strengths in high-profile platform governance approaches. In particular, our approach highlights the importance of clarifying how particular characteristics of these approaches, such as transparency, interface with the policymaking process, and with which form of legitimacy – input, throughput or output legitimacy – they should be identified. Applying this more nuanced understanding of legitimacy, in turn, can help us avoid expecting too much from individual policy proposals. For example: Multistakeholderism may improve the ability of civil society to be involved in policymaking, thus increasing throughput legitimacy, but it is not a substitute for governance *by* the people (or input legitimacy). Transparency, or a well-functioning court system, might similarly be useful in improving throughput processes, but absent democratically legitimate inputs or outputs, its overall legitimacy effects will be minimal-to-negative. And emphasizing international human rights law in and of itself will not confer legitimacy unless it results in outputs that are in line with the values of the relevant community. Failure to take these points into account will result either in platform governance policies that will not enjoy a solid grounding in (democratic) legitimacy or will impose democratically unjust policies on platform users.

Our goal in this paper is not to suggest that any single governance approach is better than the other, although it is quite clear that among these paradigmatic approaches, it is the state-based proposal that fares best in terms of democratic accountability. Rather, we argue for a more nuanced approach to the evaluation of platform-governance proposals, focusing directly on how they affect or involve

rule by the people (input) and for the people (output), in addition to the quality of policymaking (throughput). Discussions of democratic accountability must be placed front-and-centre when assessing these models. In a plurilateral world, in which people, groups and countries have good-faith disagreements about issues such as how and when to regulate speech (Haggart, 2020), it is only by considering fully the multifaceted nature of legitimacy, based in democratic accountability, will it be possible to design platform-governance models that will not only stand the test of time, but will also be accepted by the people whose lives they affect on a daily basis.

Funding

This work was supported by the Social Sciences and the Humanities Research Council of Canada.

References

- Barlow, J. P. (1996). *A declaration of the independence of cyberspace*. Electronic Frontier Foundation. <https://www EFF.org/cyberspace-independence>. (Accessed 8 December 2020).
- Bloch-Wehba, H. (2019). Global platform governance: Private power in the shadow of the state. *SMU Law Review*, 72(1), 27–80.
- von Bogdandy, A. (2004). Globalization and Europe: How to square democracy, globalization, and international law. *European Journal of International Law*, 15(5), 885–906. <https://doi.org/10.1093/ejil/chh502>
- Carr, M. (2015). Power plays in internet governance. *Millennium: Journal of International Studies*, 43(2), 640–659. <https://doi.org/10.1177/0305829814562655>
- Celeste, E. (2018). Terms of service and bills of rights: New mechanisms of constitutionalisation in the social Media environment? *International Review of Law, Computers & Technology*, 33(2), 122–138. <https://doi.org/10.1080/13600869.2018.1475898>
- Clegg, N. (2020). Welcoming the Oversight board. Facebook.com. May 6. Available at: <https://about.fb.com/news/2020/05/welcoming-the-oversight-board/>. (Accessed 8 December 2020).
- DeNardis, L. (2015). The privatization of free expression. In P. Prakash, N. Rizk, & CA Souza (Eds.), *Global Censorship: Shifting Modes, Persisting Paradigms* (pp. 11–24). New Haven: Information Society Project, Yale Law School.
- Department for Digital, Culture, Media & Sport and Home Office. (2019). *Online harms white paper*. London: Government of the United Kingdom. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf. (Accessed 15 December 2020).
- Department for Digital, Culture, Media & Sport and Home Office. (2020). *Online harms white paper: Full government response to the consultation*. London: Government of the United Kingdom. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/944310/Online_Harms_White_Paper_Full_Government_Response_to_the_consultation_CP_354_CCS001_CCS1220695430-001_V2.pdf. (Accessed 15 December 2020).
- Department for Digital, Culture, Media & Sport and Home Office. (2020a). *Online harms white paper - initial consultation response*. London: Government of the United Kingdom. <https://www.gov.uk/government/consultations/online-harms-white-paper/public-feedback/online-harms-white-paper-initial-consultation-response>. (Accessed 15 December 2020).
- Di Gregorio, G. (2020). The rise of digital constitutionalism in Europe. *International Journal of Constitutional Law*. In Press.
- Douek, E. (2019). Facebook's "Oversight Board": Move fast with stable infrastructure and humility. *North Carolina Journal of Law and Technology*, 21(1), 1–78.
- DeNardis, L. (2012). Hidden Levers of Internet Control: An infrastructure-based theory of internet governance. *Information, Communication & Society*, 5, 720–738. <https://doi.org/10.1080/1369118X.2012.659199>
- Facebook. (2020). *Oversight Board Bylaws*. Facebook.com. https://about.fb.com/wp-content/uploads/2020/01/Bylaws_v6.pdf. (Accessed 6 March 2020).
- Gasser, U. (2006). Regulating search engines: Taking stock and looking ahead. *Yale Journal of Law & Technology*, 8(1), 202–234.
- Giddens, A. (1984). *The Constitution of Society*. Berkeley, CA: University of California Press.
- Gorwa, R. (2019). What is platform governance? *Information, Communication & Society*, 22(6), 854–871. <https://doi.org/10.1080/1369118X.2019.1573914>
- Gorwa, R. (2019a). The platform governance triangle: Conceptualising the informal regulation of online content. *Internet Policy Review*, 8(2), 1–22. <https://doi.org/10.14763/2019.2.1407>
- Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, 7(1), 1–15. <https://doi.org/10.1177/2053951719897945>
- Hartmann, I. (2017). Let the users be the filter? Crowdsourced filtering to avoid online intermediary liability. *Journal of the Oxford Centre for Socio-Legal Studies*, 2017(1), 21–47. <https://joxs.files.wordpress.com/2017/11/issue-1-2017-let-the-user-be-the-filter.pdf>
- Haggart, B. (2020). Global platform governance and the internet-governance impossibility theorem. *Journal of Digital Media & Policy*, 11(3), 229–321. <https://doi.org/10.1386/jdmp.00028.1>
- Jørgensen, R. F. (2017). What platforms mean when they talk about human rights. *Policy & Internet*, 9(3), 280–296. <https://doi.org/10.1002/poi3.152>
- Johnson, D. R., & Post, D. (1996). Law and borders — The rise of law in cyberspace. *Stanford Law Review*, 485, 1367–1402. <https://doi.org/10.2307/1229390>
- Kaye, D. (2018). *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. Report no. A/HRC/38/35. Geneva: United Nations, Human Rights Council https://ap.ohchr.org/documents/dpage_e.aspx?si=A/HRC/38/35. (Accessed 8 December 2020).
- Kaye, D. (2019). *Speech Police: The GobaL Struggle to Govern the Internet*. New York: Columbia Global Reports.
- Kelion, L. (2020). *Online harms law to let regulator block apps in UK, 15 December*. BBC News. <https://www.bbc.com/news/technology-55302431>. (Accessed 15 December 2020).
- Kettemann, M. C. (2020). *The Normative Order of the Internet: A Theory of Rule and Regulation Online*. Oxford: Oxford University Press.
- Kettemann, M. C., & Schulz, W. (2020). *Setting rules for 2.7 billion. A (first) Look into Facebook's norm-making system: Results of a pilot study*. Hamburg: Working papers of the hans-bredow-institut, works in progress # 1. January. <https://doi.org/10.21241/ssor.71724>. Available at: https://www.hans-bredow-institut.de/uploads/media/default/cms/media/k0gixdi_AP_WiP001InsideFacebook.pdf. (Accessed 8 December 2020).
- Keller, C. I. (2020). Policy by judicialisation: the institutional framework for intermediary liability in Brazil. *International Review of Law, Computers & Technology*, 1–19. <https://doi.org/10.1080/13600869.2020.1792035>
- Klonick, K. (2017). The new governors: The people, rules, and processes governing online speech. *Harvard Law Review* 131: 1598–1670. Available at: https://harvardlawreview.org/wp-content/uploads/2018/04/1598-1670_Online.pdf
- Klonick, K. (2020). The Facebook Oversight Board: Creating an independent institution to adjudicate online free expression. *The Yale Law Journal*, 129, 2418–2499.
- Kumm, M. (2004). The legitimacy of international law: A constitutionalist framework of analysis. *European Journal of International Law*, 15(5), 907–931. <https://doi.org/10.1093/ejil/chh503>
- Lara, J. C., & Vera, F. (2013). "La responsabilidad de los prestadores de servicios de internet." Policy Paper n. 3. ONG Derechos Digitales. <https://www.derechosdigitales.org/wp-content/uploads/pp03.pdf>. (Accessed 8 December 2020).
- Lessig, L. (2006). *Code: Version 2.0*. New York: Basic Books.
- Machado, J. E. M. (2001). *Liberdade de Expressão. Dimensões Constitucionais da Esfera Pública no Sistema Social*. Coimbra: Coimbra Editora.
- Manila Principles. (2015). On intermediary liability. <https://www EFF.org/files/2015/10/31>. (Accessed 6 March 2020).
- Manila Principles on Intermediary Liability. (2015). Background document. https://www EFF.org/files/2015/07/08/manila_principles_background_paper.pdf. (Accessed 6 March 2020).
- Murdoch, Z., Connolly, S., & Kassim, H. (2018). Administrative legitimacy and the democratic deficit of the European Union. *J. Eur. Public Policy*, 253, 389–408. <https://doi.org/10.1080/13501763.2016.1268193>

- Announcing the Oversight Board's first case decisions, (2021–). (Accessed 15 April 2021).
- The Oversight Board is accepting user appeals to remove content from Facebook and Instagram, (2021–). (Accessed 15 April 2021).
- Padovani, C., & Santaniello, M. (2018). Digital constitutionalism: Fundamental rights and power limitation in the Internet eco-system. *International Communication Gazette*, 80(4), 295–301. <https://doi.org/10.1177/1748048518757114>
- Pereira, J. (2016). As garantias constitucionais entre utilidade e substância: uma crítica ao uso de argumentos pragmatistas em desfavor dos direitos fundamentais. *Direitos Fundamentais & Justiça, Belo Horizonte*, 10(35), 345–373.
- Pernice, I. (2015). *Global constitutionalism and the internet. Taking people seriously*. Berlin: Humboldt Institute for Internet and Society. Discussion Paper No. 2015-01. <https://www.hiig.de/en/publication/global-constitutionalism-and-the-internet-taking-people-seriously-2/>. (Accessed 8 December 2020)
- Pohle, J. (2016). Multistakeholder governance process as production sites: Enhanced cooperation “in the making”. *Internet Policy Review*, 5(3), 1–19. <https://doi.org/10.14763/2016.3.432>
- Powers, S., & Jablonski, M. (2015). *The Real Cyber War: The Political Economy of Internet Freedom*. Chicago: University of Chicago Press.
- Redeker, D., Gill, L., & Gasser, U. (2015). Towards digital constitutionalism? Mapping attempts to craft an internet bill of rights. *International Communication Gazette*, 80(4), 302–319. <https://doi.org/10.1177/1748048518757121>
- Roberts, S. T. (2019). *Behind the screen: Content moderation in the age of social media*. New Haven, CT: Yale University Press.
- Rodrik, D. (2011). *The globalization paradox: Democracy and the future of the world economy*. New York: W.W. Norton & Company.
- Reidenberg, J. R. (1996). Governing networks and rule-making in cyberspace. *Emory Law Journal*, 45, 912–930. Available at: https://ir.lawnet.fordham.edu/faculty_scholarship/29.
- de la Rue, Frank. (2013). *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. A/HRC/23/40. Geneva: United Nations, Human Rights Council. <https://undocs.org/A/HRC/23/40>. (Accessed 6 March 2020).
- Scharpf, F. W. (1996). Democratic policy in Europe. *European Law Journal*, 2(2), 136–155. <https://doi.org/10.1111/j.1468-0386.1996.tb00022.x>
- Scharpf, F. W. (1999). *Governing in Europe: Effective and Democratic?* Oxford: Oxford University Press.
- Schmidt, V. A. (2013). Democracy and legitimacy in the European Union revisited: Input, output and “throughput”. *Political Studies*, 61(1), 2–22. <https://doi.org/10.1111/j.1467-9248.2012.00962.x>
- Stivers, C. (2018). “You don’t know what you’ve got till it’s gone”: Rejoinder to Aaron Wachhaus’s article, “Platform governance”. *Administrative Theory and Praxis*, 40(2), 150–158. <https://doi.org/10.1080/10841806.2017.1420745>
- Suzor, N. (2018). Digital constitutionalism: Using the rule of law to evaluate the legitimacy of governance by platforms. *Social Media + Society*, 4(3), 1–11. <https://doi.org/10.1177/2056305118787812>
- Suzor, N. (2019). *Lawless: The Secret Rules that Govern our Lives*. Oxford: Oxford University Press.
- Suzor, N., Van, G. T., & West, S. M. W. (2018). Evaluating the legitimacy of platform governance: A review of research and a shared research agenda. *International Communication Gazette*, 80(4), 385–400. <https://doi.org/10.1177/1748048518757142>
- Sylvain, O. (2010). Internet governance and democratic legitimacy. *Federal Communications Law Journal*, 62(2), 205–273.
- Tufekci, Z. (2018). It’s the (Democracy-Poisoning) golden age of free speech. *Wired*. January 16 <https://www.wired.com/story/free-speech-issue-tech-turmoil-new-censorship/>. (Accessed 8 December 2020).
- Weber, M. (1919). Politics as a vocation (Translated and edited) (1946). In H. H. Gerth, & C. Wright Mills (Eds.), *From Max Weber: Essays in Sociology* (pp. 77–128). New York: Oxford University Press <http://fs2.american.edu/dfagel/www/Class%20Readings/Weber/PoliticsAsAVocation.pdf>.
- Wu, T. (2010). Is internet exceptionalism dead? In B. Szoka, & A. Marcus (Eds.), *The Next Digital Decade: Essays on the Future of the Internet*. Washington: TechFreedom.
- Xu, X., & Hager, M. (2020). First Nations urge B.C. privacy watchdog to disclose localized COVID-19 data. In *The Globe and Mail*. September 15. <https://www.theglobeandmail.com/canada/british-columbia/article-first-nations-urge-bc-to-share-localized-covid-19-data/>. (Accessed 8 December 2020).