Democratic legitimacy in global platform governance

Blayne Haggarts, Brock University
Clara Iglesias Keller, WZB Berlin Social Science Center

Abstract

The goal of this paper is to propose a democratic legitimacy framework in order to clarify the different forms of legitimacy in platform-governance proposals and policies, and in doing so clarify terms of debate and allow 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. We describe 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 (rule-of-law-focused regimes); the human-rights-focused framework proposed by then UN free speech Special Rapporteur David Kaye; and the United Kingdom's Online Harms White Paper (domestic regime). The paper concludes with some final thoughts about the implications of our analysis for future policymaking.

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). Academic work has similarly focused on questions of legitimacy in platform governance (e.g., Suzor, Geelen and West, 2018; Suzor 2019; Kaye 2019), and can be seen as a continuation of a long-standing debate around gov ernance 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, p. 723) to the policy decisions regarding speech made or enforced by content intermediaries (p. 725; see also Post and Johnson 1996; Reidenberg 1998; Lessig 2006). More recently, proposals intending to encourage greater legitimacy in platform regulation have appealed to rule-of law principles (Suzor 2019; Manila Principles), as well as to human rights laws and norms (Kaye 2019).

While appeals to ground platform regulation in human-rights laws and the rule of law are commendable, these represent only partial contributions to a regulatory framework's legitimacy and thus to its continued success.

This paper argues that insufficient attention has been paid to the issue of democratic legitimacy. Furthermore, rather than treat legitimacy as a one-dimensional concept (based, for example on adherence to human-rights 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 democratic legitimacy as it relates to the European Union. According to this framework, the 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 highlight 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 regimes and proposals allows us to highlight their legitimacy-related strengths and weaknesses in a more nuanced way, while also offering a clearer conceptual framework for the development of future proposals. In particular, it highlights the extent to which proposals that are not rooted in domestic-democratic legitimacy tend to overemphasize throughput legitimacy to the detriment of input and output legitimacy. They also tend to sideline state regulation (including by democratic states), possibly as a consequence of their 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 (rule-of-law-focused regimes); the human-rights-focused framework proposed by then UN free speech Special Rapporteur David Kaye; and the United Kingdom's Online Harms White Paper (domestic regime). The fourth section offers our analysis. The paper concludes with some final thoughts about the implications of our analysis for future policymaking.

1. Democratic legitimacy in internet and platform governance

Defining legitimacy

We argue that democratic legitimacy is the appropriate starting point for assessing the legitimacy of platform governance regimes and proposals. Like other international and supranational institutions, online platforms are characterized by a democratic deficit, in this case due to their status as privately controlled entities whose internally set norms directly influence the behavior of billions of users who have little-to-no say over their construction (Reidenberg 1998; Lessig, 2006).

Within the nation-state, a regime'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" (Bogdandy 2004, p. 887). This notion is centred 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" (UNDHR 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, 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, 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, 27).

These different references for nation-state legitimacy – democratic participation, individual constriction, 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 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. David Post and David Johnson (1996), for example, argued against government control of the internet and in favour 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, 180), the challenge of finding a source of legitimacy outside the reach of majoritarian institutions of electoral representation (Scharpf, 1996; Schmidt, 2012) 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, 5). Multistakeholderism as a governance approach is rooted in the idea that the policy 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, 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 (Gillepsie, 2017), both led by workers (Roberts, 2018) and automated tools (Gorwa, Binns, and 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" (Schulz and Kettemann, 2019).

The question of how to address these legitimacy problems has been approached from different angles. In an attempt to bring an approximation of democratic participation to platform rule-making, Ivar Hartmann, e.g., advocates for greater user engagement in content moderation, which would fulfill what they see as "Internet's deep-rooted wish for self-regulation" (2019, 1). This is in line with internet exceptionalism proponents such as John Perry Barlow, 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, 154) argues that "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".

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). The latter's work 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, 387; 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" (Gil, Redeker, Gasser, 2015, p. 2), in order to ground internet governance in "fundamental rights and democratic principles" (Padovani and Santaniello 2018; see also Gil, Redeker, Gasser, 2015; Celeste 2018; and De Gregorio 2020).

Although they provide a comprehensive list of high-level human-values (Suzor, Geelen, and West 2018, 388–89), their analysis fails to come to terms with the reality that some interpretations of "freedom of expression and privacy rights" can often come into conflict with the right to be free from discrimination. It does not fully consider that the extent to which a country or company adheres to "human rights values" is often in the eye of the beholder (Noble 2018; Eubanks 2018). As well, while the questions of values and process as legitimizing forces are discussed extensively, the issue of which actors possess rule-setting legitimacy 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-evaluation projects; in their 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 UNDHR's Article 21 (Suzor, Geelen, and West 2018, 392). Most importantly, by taking global platforms' current private-ordering regime as a given rather than something that can be changed through regulation, it leaves no room to consider even a theoretical role for the state as regulator.

As this section shows, scholars of internet governance and platform governance have been interested in legitimacy issues related to participation, rule-making and user engagement. At the same time, however, they continue to struggle with the problem of democratic legitimacy. In the next section, we outline a more nuanced theoretical framework that can more comprehensively assess platform governance legitimacy claims.

2. Assessing democratic legitimacy outside the nation-state

Schmidt (2013), building on Scharpf (1970), proposes a three-pronged framework for evaluating the EU's democratic legitimacy. 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," while throughput legitimacy (the concept of which is Schmidt's novel contribution to this literature) focuses on "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, 2).

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 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, 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, 7).

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 representation based in a citizenry. 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 possess legitimacy not just because they possess a technical expertise or are isolated from politics. Rather, it is "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 – meaning that their non-majoritarian throughput processes are balanced by majoritarian input politics." (Schmidt 2013, p. 10). Absent this power to alter a platform by the citizens affected by the platform, there can be no input legitimacy.

Output legitimacy

Output legitimacy, meanwhile, covers "the problem-solving quality of the laws and rules" (Schmidt 2013, 4), or "effectiveness of … policy outcomes *for* the people" (Schmidt 2013, 2). In considering output legitimacy, EU scholars consider not only "the community-enhancing performance of the policies themselves" but also the extent to which these policies are communicatively legitimated through discursive and communicative actions (Schmidt 2013, 5). They are policies that "work effectively while resonating with citizens' values and identities" (Schmidt 2013, 7). To take an example, it is not enough to say, for example, that, say, "human rights" resonates with people. Rather it is whether the interpretation and operationalization of human rights resonate with a particular community's values, and whether the outcomes are seen as just.

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, 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, 6). It "demands institutional and constructive governance processes that work with efficacy, accountability, transparency, inclusiveness and openness" (Schmidt 2013).

The mechanisms of throughput legitimacy, meanwhile, are "their efficacy, accountability, transparency, inclusiveness and openness to interest intermediation" (Schmidt 2013, 6). By efficacy, Schmidt means how well they are able to actually do their jobs. 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, 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, 6). Finally, inclusiveness and openness of institutional processes to civil society participation is legitimizing as a balance to "access and influence among organized interests representing business" (Schmidt 2013, 6–7). Since the major platforms are themselves businesses, in the context of platform governance it is the platform itself that these groups (which would also include private-sector actors affected by the platform in question) would be balancing against. "Interest-based throughput" differs from "interest group-based 'input'," in that the latter influences (representative) rule-makers, while in the former, interest groups are themselves "an integral part of the very throughput process of … governance" (Schmidt 2013, 7). This distinction is particularly useful when considering calls for greater multistakeholder participation in platform governance. From this perspective, multistakeholder processes should be seen primarily as part of the throughput process of governance, of creating the rules.

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, 3). For example, one can imagine general acceptance of a human rights-driven content-moderation decision even if it goes against one's values, if you could argue it was arrived at democratically.

When throughput is added to the mix, however, the relationships become more complicated. hile 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, 9). Governments rarely receive credit for a process that runs smoothly because when they do, they're seen as merely doing what they should do anyway. When the process goes awry, voters remember and can judge harshly.

This finding is particularly important for our understanding of platform-governance legitimacy, especially given the emphasis so many proposals and activists place on transparency (a throughput issue).

A framework for evaluating platform governance

Based on the preceding discussion, we propose the following steps to evaluating the legitimacy of actual and proposed forms of platform governance, along with examples of what types of activities fall under each.

i. Input legitimacy
- The degree to which the people affected by a platform can participate in *all* rules governing the platform.
- The degree to which the people running the platform are responsive to citizen concerns emerging from this participation.

Input legitimacy for platforms thus focuses on the ability of citizens to control the very form of the platform itself, as well as what it does and how it does it. It is reflected in, but not reducible to, the extent to which platforms are responsive to citizens' demands.

ii. Output legitimacy
- The extent to which a platform's actions enhance the lives of the people affected by it.
- The extent to which these actions fit the values and identities of the people affected by it.

Output legitimacy, then, looks at what a platform does. The extent to which it promotes human rights (in general or specific ones), for example, would be considered under this category.

iii. Throughput legitimacy
- The quality of governance processes. These include:
  - Efficacy – Does the process actually achieve its intended outcome?
  - Accountability—Are actors held accountable for decisions and responsive to "participatory input demands?
  - Transparency—Do people have enough information about decision-making processes and rules to hold rule-makers accountable?
    - Corollary: Is there an institution within which transparency-based information can provide accountability?
  - Inclusiveness and openness to civil-society participation—Do non-platform actors have an ongoing role in setting platform policy?

In this framing, multistakeholder governance, which is often seen as a form of, or replacement for, democratically accountable input legitimacy (e.g., Suzor, Geelen, and West 2018, 389), is more properly classified as being a throughput in which outside groups are involved in policymaking.

iv. Interaction effects and overall evaluation
- 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.
- Any platform-governance legitimacy assessment must take into consideration the likely interaction effects among the three forms of legitimacy.
  - 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 can negate high throughput legitimacy.
  - High throughput legitimacy tends to have a minimal effect on input and/or output legitimacy.
  - Low throughput legitimacy tends to have a negative effect on input and/or output legitimacy.

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

3. Four paradigmatic case studies

3.1. Facebook's Oversight Board

Since 2019, Facebook has been in the process of implementing its "Oversight Board", which will serve as an independent appeal body with competency to review Facebook's content moderation decisions. Its goal is "to protect free expression by making principled, independent decisions about important pieces of content and by issuing policy advisory opinions on Facebook's content policies" (Bylaws, p. 5). It will deliberate on content disputes according to the website's Community Standards.

The Oversight Board will only analyze appeals regarding content that was removed based on these Standards - i.e., Facebook's own rules. 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.

Within this already restricted scope, the possibility for appeals regarding content taken down according to the Community Standards represents "a sheer volume of content moderation decisions that Facebook makes every day", as it "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).

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.

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 *self-regulation*.

From Klonick's description of the Board's conception one can perceive the intention to promote extensive participation in this process, possibly as an attempt of improving input legitimacy (2020, p. 2451). The author 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). In the face of the 2.5 billion

Facebook users, this rate of participation does not exactly represent high standards for input legitimacy.

Further, neither this consultation process nor the structure of the Oversight Board improved 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.

Throughput legitimacy

Throughput legitimacy offers the strongest legitimacy claim of the Oversight Board, because it focuses on inserting transparency, due process, and accountability into content moderation decisions. It purports to provide a solution to Facebook's current "unchecked system" for users' speech governance (Klonick, 2020, p. 2476 and Douek, 2020, p. 46). The degree to which this check will indeed be independent is, as of now, questionable. Even though Facebook has implemented a few mechanisms that would assure independence, a conclusive assessment of this feature can only be achieved once the Board is operational.

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, the Bylaws commit the Board to making all case decisions publicly available, archiving them in a database and publishingannual 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 under another instance's scrutiny. Nevertheless, there is a strong claim for due process and procedural legitimacy.

Due process is in fact perceived as one of its defining characteristics (Douek, 2019, p. 6). The central goal of this structure is to improve it and grant Facebook users the possibility of having their content controversy examined by a selection of people from different geographical origins, notorious expertise and an alleged independence from Facebook.

As much as throughput legitimacy is the strongest claim, it is not a completely fulfilled one. Openness to interests' intermediation, for instance, is restricted to the possibility of appeal. There is no provision for other sorts of participation mechanisms whatsoever.

Lastly, to assess the Board's efficacy - will it be able to function? - would depend on what one understands as "functioning". Given its concise terms, the Bylaws do not offer much insight on that, since the only reference to objectives comes in the Introduction: "[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". Procedurally, that seems like a feasible goal. However, in substantive terms—that is, if freedom of expression will indeed be better protected after the Board—then the output legitimacy is not so promising.

Instituting an effective and throughput-legitimate framework—especially when its scope is a restricted one—will depend on how this operation is in fact implemented: i.e., if the process is indeed a thorough independent review, or a "way of ensuring the effective functioning of a bureaucratic system and rule enforcement by creating a mechanism for error correction" (Douek, 2019, p. 6).

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.

Besides the already described restricted scope, Board decisions risk being undermined by the 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, 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, 2475). Even though the Board's decisions provide users with an extra, expertise-driven level of scrutiny, they are still likely to clash with their norms and expectations (Klonick, 2020, 2475).

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 a meaningful transparency and reasoning process underlying specific content moderation decisions (Klonick, 2019). At the end of the day, this is an instrumental mechanism to assure procedural guarantees to a small fraction of online content disputes. It does not have input legitimacy, nor are its results guaranteed; its biggest claim for legitimacy therefore is that fact that it exists in this form and abides by pre-established procedures.

The Board does not propose a new normative framework that could exercise a significant influence on output; the Community Standards and Values that 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 to decision making processes or to the norms' construction and review process (regardless of the global consultation process).

A final assessment of the Board can only be undertaken once it is fully operational and has ruled on some cases. In terms of legitimacy, the main concern is that robust throughput may not be enough to compensate for the lack of input and output legitimacy. That said, 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.

3.2. Judicial adjudication

Overall, court adjudication provides legal certainty to the interpretation of the law in concrete cases, and 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), aiming at developing "intermediary liability policies that can foster and protect a free and open Internet" (Manila Principles, Background Document). Court adjudication is the second principle, according to which "[c]ontent must not be required to be restricted without an order by a judicial authority" (Manila Principles). 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).

It has also been identified as a legitimacy standard of intermediary liability by UN Specialist reports (Kaye, 2018; Rue, 2013) and academic literature (Urs and Schulz, 2015). Some countries embraced it as a mandatory requirement for liability, which is considered a democratic practice (Gasser and Schulz, 2015) for its alleged privilege of freedom of expression (Keller, 2020). The assumption here 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), reducing platforms' incentives to overblocking (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 legitimate content.

Beyond the Manila Principles, this type of regime has been adopted by Brazil in its Marco Civil. Under section 19, intermediaries can only be liable for third-party harmful content if they fail to remove content after receiving judicial take-down notification. Instead of being potentially responsible for any committed (or acknowledged) infringement, findings of liability rests 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 and 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.

Input legitimacy

At first glance, input legitimacy can be considered the strongest claim for submitting online content disputes to 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*.But 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 input legitimacy. Adjudication of intermediary liability can generate "policy like" effects on the online content governance framework, bypassing the government branches that are initially responsible for implementing policies towards speech (Keller 2020, 16-17). Thus, there is a claim to be made, therefore, for a deficit in their input legitimacy to adjudicate these matters (Pereira, 2015). Constitutional law literature has already recognized that the judicial branch is not necessarily reflective of the flow of popular representation that basis public policy choices (Hirschl 2008).

In this sense, their legitimacy for adjudicating cases that have impacts in such choices can be questioned (Pereira 2015). Because freedom of expression is both an individual right, but also a collective guarantee (Machado 2002), its promotion also depends on positive state provisions that are inherently under the legislative and original executive competences.

Throughput legitimacy

Initially, adjudication would 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.

In contrast, though, the qualities of openness and efficacy 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 the one brought by parties. It should be noticed that this can be mitigated in systems that allow for *amicus curiae* in certain lawsuits[1]. 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.

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' lack expertise to adjudicate matters that require specific technical knowledge, plus the lawsuit scope limitations mentioned above, can

---

[1] "Amicus curiae, (Latin: "friend of the court"), one who assists the court by furnishing information or advice regarding questions of law or fact. He is not a party to a lawsuit and thus differs from an intervenor, who has a direct interest in the outcome of the lawsuit and is therefore permitted to participate as a party to the suit." (Britannica Academic 1998).

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 [2]. 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 reality is that a significant proportion of online content regulation will effectively be left to the discretion of intermediaries who are then free to apply their own product's policies in matters of content moderation, according to their own criteria.

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.

That so, adjudication, particularly in a democracy, can play an important role in protecting individual rights; similarly, the intermediary liability standard discussed in this section is a defensible policy choice on its own. However, it is important to stress 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.

3.3. A human-rights-centric framework

---

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

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[es] 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 framework" 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.

Input legitimacy

Democratic input legitimacy is determined by who sets the rules. In this case, the rules refer to how human rights standards by and affecting platforms are to be interpreted or operationalized. Kaye implicitly recognizes that human rights standards are highly contestable (Kaye 2018, paragraph 41). However, his overall proposal tends to downplay these challenges, treating them as technical exercises with one obvious answer (promote "human rights"), rather than political/cultural exercises in which the extent to which specific laws and rules achieve these objectives are always contestable. This elision presents problems for output legitimacy, discussed below.

The application of international human rights laws and norms is a matter of interpretation. Consequently, the question of who interprets and operationalizes these international laws and norms is the central governance issue. Kayes' framework pays little attention to this interpreter role. Although Kaye, in this document and the resulting book (Kaye 2019), clearly recognizes the problems that arise when companies are left to set their own rules, his framework fails to challenge their supremacy. Both users and civil society perhaps in the form of a hypothetical consultative Social Media Council (Docquir 2019; Tworek 2019) – are treated as advisors without decisive rule-making, or agenda-setting powers, let alone the ability to remove the decision-makers running the platform. 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).

Companies, in Kaye's framework, remain the ultimate deciders of how "human rights" are constituted.

Nor does Kaye consider the fundamental question of civil society's or the user community's representativeness. Even the smallest society contains divergent views. 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.

Of course, there exists a process and actor for adjudicating such questions: 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, the platform.

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

Throughput legitimacy

In the absence of input legitimacy, Kaye's framework is rooted primarily in throughput legitimacy and a flawed understanding of 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. 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, paragrah 54). The key word, again, is consultation: companies remain in charge of the process.

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) 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 a far more universal view of (international) "human rights law" against which national exceptions would be applied.

That reasonable people can disagree about where to draw this line highlights the fundamental legitimacy flaw in Kaye's proposal: it assumes that the regulatory question is, "Where should we draw the line?" In reality, the challenge is, "Who should be responsible for drawing the line?"

Interactions and overall assessment

This assessment suggests that Kaye's human rights-focused framework does not draw its 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 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, Kayes framework focuses to a significant on throughput legitimacy particularly as it relates to transparency. (Actual accountability, meanwhile, is strangely 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 Facebook is more open about how it came to make unpopular decisions is unlikely to win it any support among the people it rules.

3.4. United Kingdom Online Harms White Paper

The United Kingdom's Online Harms White Paper was released in April 2019. It proposes a domestic regulatory framework to address online "illegal and unacceptable content and activity" (Secretary of State, Digital, Culture, Media & Sport and Secretary of State, Home Department 2019, paragraph 2). The White Paper proposes the establishment of a "statutory duty of care" for online companies similar to the "high-level policy commitments" proposed by Kaye. This policy would be "overseen and enforced by an independent regulator" (2019, paragraph 17); the regulator would be responsible for setting out how companies will fulfill this new duty (2019, paragraph 20). The government would also have "the power to direct the regulator in relation to codes of practice on terrorist activity or child sexual exploitation and abuse … signed off by the Home Secretary" (2019, paragraph 21).

Of particular interest is the fact that this independent regulator would be designed to address both "illegal" and "harmful" content that is not necessarily illegal (see table 1). While the definition of "illegal content" is determined by statute, the definition of harmful content that is not illegal (which the White Paper refers to as "legal harms" (2019, paragraph 2.8)) would be set out in codes created by the independent regulator (2019, paragraph 3.6). Companies would be required to show that they have adopted terms and conditions consistent with this duty of care (2019, paragraph 18).

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).
Organised immigration crime.	Coercive behaviour.	
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.		
Sexting of indecent images by under 18s (creating, possessing, copying or distributing indecent or sexual images of children and young people under the age of 18).		

The regulator would also "have the power to require annual transparency reports from companies in scope" – to be published online "outlining the prevalence of harmful content on their platforms and what counter measures they are taking to address these" (2019, paragraph 23). It would also have the power "to require additional information, including about the impact of algorithms in selecting content for users and to ensure that companies proactively report on both emerging and known harms" (2019, paragraph 23), and to encourage independent

researchers' access to companies' data (2019, paragraph 24). 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 ((2019, paragraphs 25-26).

While the government has committed to implementing the White Paper, as of August 2020, no legislation has yet been tabled.

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 (Secretary of State, Department for Digital, Culture, Media & Sport and Secretary of State, Home Office 2020, paragraphs 2 and 10; Secretary of State, Digital, Culture, Media & Sport and Secretary of State, Home Department 2019, 97–98). Industry, organizations and individuals were represented in both phases, as detailed in the government's *Initial Consultation Response*.

With respect to ongoing rule-making, the "duty of care" will be set out in parliamentary statute, as are illegal harms (e.g., child pornography) that will fall under the regulator's purview. That said, the White Paper remarks that "Parliament's role in relation to codes of practice and guidance issued by other regulators varies across different regulatory regimes, ranging from formal approval to no specific role. We will consider options for the role of Parliament as we develop these proposals in more detail" (Secretary of State, Digital, Culture, Media & Sport and Secretary of State, Home Department 2019, paragraph 3.32). The regulator will have the power to define and enforce the much more nebulous "legal harms" category. In other words, these consequential definitions would be set in "the shadow of politics," but not more directly via Parliament.

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 pays significant attention to transparency as a way to improve regulatory quality and platform-user understanding (2019, paragraphs 3.13-3.25). It also claims that the regulator will adopt "an evidence-based approach to regulatory activity" (2019, paragraph 5.13) 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, paragraph 3.19). With respect to ensuring its work is based on research, the government expects "the regulator will continue to support research in this area to inform future action and, if necessary, set clear expectations for companies to prevent harm to their users" (2019, box 15). These moves suggest at least a nod toward inclusiveness in the policy-setting process.

Like other independent regulators, such as central banks, and as already noted, this independent regulator would not be directly accountable to the citizenry, although like such agencies 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.

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" (ICR paragraph 5). 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.

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. Most interesting is how leaving the definition of "legal harms" to an independent regulator creates a conundrum. The isolation from popular influence (via Parliament and the government) on the one hand could be seen to reduce its democratic legitimacy. On the other hand, this isolation from popular pressure, in a democratically created institution embedded within a democratic regime, could also be seen as increasing the regulator's throughput legitimacy, ensuring that free-speech issues are not unduly influenced by politicians, a long-standing concern of free-speech advocates. That said, these same advocates are likely to be categorically opposed to any regulatory authority definitions of "legal but harmful" actions. The regulator's output legitimacy in this area is likely to be a function of the extent to which it is seen to reflect citizens' perceptions about such actions. That there is likely to be a diverse spectrum of opinions on this issue highlights the importance of the democratic nature of the policy-making process (input legitimacy) in shaping overall legitimacy.

4. 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.

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.

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 scrules. The Boards structure does not promote any change or enhances 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 draw their input legitimacy from the "shadow of politics." However, absent a democratic polity, such courts cannot be said to have, on their own, democratic input legitimacy. 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."

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 linked to the state in its area of expertise (harmful but not illegal content) is obvious. More interestingly, however, is the relative neglect of the role of the state in the Kaye and adjudication cases. 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; the legitimacy of court decisions depends on the policies that they are charged with implementing. Furthermore, as Keller (20120) courts alone cannot make policy, as they lack the expertise and capacity to do so.

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 between democratic and non-democratic states when it comes to the legitimacy of their regulatory efforts. As a UN Special Rapporteur, it perhaps might have been impolitic to make this distinction, but from an input-legitimacy perspective, it is hard to justify effectively placing on the same level the regulatory efforts of a democratic country like New Zealand with those of an authoritarian dictatorship like China. In doing so, Kaye's proposal, alongside the domestic-judicial emphasis of adjudication processes and Facebook's Oversight Board, undermines the notion of democratic-state platform regulation.

Output legitimacy and 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 regimes. 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 Advisory 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 underplayed.

5. Conclusion

By assessing how four paradigmatic platform-governance proposals/policies fared in terms of three types of policy legitimacy – input, output, and throughput – this paper

highlighted the need for a more nuanced view of legitimacy when designing and assessing platform-governance models. It further argues for the need to privilege democratic accountability in this work. 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.

References

"Amicus curiae" Britannica Academic, Encyclopædia Britannica, 20 Jul. 1998. academic.eb.com/levels/collegiate/article/amicus-curiae/6171#. Accessed 9 Sep. 2020.

Bogdandy, Armin von (2004). "Globalization and Europe: How to Square Democracy, Globalization, and International Law", The European Journal of International Law Vol. 15 n. 5.

Carr, Madeline (2015). "Power Plays in Internet Governance." Millennium: Journal of International Studies Vol 43, n. 2. pp. 640-6509.

Celeste, Edoardo (2018)."Terms of Service and Bills of Rights:New Mechanisms of Constitutionalisation in the Social Media Environment?" Accessed March 6, 2020.

Clegg, Nick (2020). "Welcoming the Oversight Board." Facebook.com. May 6. Available at: https://about.fb.com/news/2020/05/welcoming-the-oversight-board/.

DeNardis, Laura (2015). The Privatization of Free Expression. In: Pranesh Prakash, Nagla Rizk and Carlos Affonso Souza. "Global Censorship – Shifting Modes, Persisting Paradigms". Information Society Project, Yale Law School, p. 11-24.

Douek, Evelyn (2019). "Facebook's "Oversight Board": move fast with stable infrastructure and humility", North Carolina Journal of Law & Technology, volume 21, issue 1.

Gasser, Urs (2006). "Regulating search engines: taking stock and looking ahead", 9 Yale Journal of Law & Technology, 124.

Giddens, Anthony (1984). *The Constitution of Society*. Berkeley, CA: University of California Press.

Gill, L., Redeker, Dennis and Gasser, Urs. 2015. "Towards Digital Constitutionalism? Mapping Attempts to Craft an Internet Bill of Rights." Berkman Klein Center for Internet & Society Research Publication No. 2015-15.

Gorwa, Robert, Binns, Reuben and Katzenbach, Christian (2020). "Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance." *Big Data & Society*. doi:10.1177/2053951719897945.

Gregorio, Giovanni Di (2020). "The rise of digital constitutionalism in Europe", International Journal of Constitutional Law. Forthcoming.

Haggart, Blayne (2019). "Global platform governance and the internet-governance impossibility theorem." Presented at GigaNet19 Symposium, Berlin, Germany, November 25.

Hartmann, Ivar (2014). "Let The Users Be The Filter? Crowdsourced Filtering To Avoid Online Intermediary Liability", Journal of the Oxford Centre for Socio-Legal Studies, v. 2017, issue 1, p. 21-47.

Keller, Clara Iglesias (2019). Policy by judicialisation: the institutional framework for intermediary liability in Brazil, Presented at GigaNet19 Symposium, Berlin, Germany, November 25.

Kettemann, Matthias C.; Schulz, Wolfgang: Setting Rules for 2.7 Billion. A (First) Look into Fa- cebook's Norm-Making System: Results of a Pilot Study. Hamburg: Working Papers of the Hans-Bredow-Institut | Works in Progress # 1, January 2020.

Klonick, Kate (2020). "The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression", Yale Law Journal, Vol. 129, n. 2418.

Kumm, Matthias (2004). "The Legitimacy of International Law: A Constitutionalist Framework of Analysis", The European Journal of International Law, vol. 15, n. 5.

Lara, Juan Carlos and Vera, Francisco (2013). "La responsabilidad de los prestadores de servicios de internet", Policy Paper n. 3, ONG Derechos Digitales.

Lindsay, David (2013. "What do the .XXX disputes tell us about Internet governance? ICANN's legitimacy deficit in context", Telecommunications Journal of Australia, Volume 63, Number 3. DOI: //dx.doi.org/10.7790/tja.v63i3.432.

Machado, Jónathas (2002). "Liberdade de Expressão. Dimensões Constitucionais da Esfera Pública no Sistema Social". Coimbra: Coimbra Editora.

Manila Principles on Intermediary Liability."Best Practices Guidelines for Limiting Intermediary Liability for Content to Promote Freedom of Expression and Innovation," v. 1.0. 2015. Accessed March 6, 2020. https://www.eff.org/files/2015/10/31.

Padovani, Claudia and Santaniello, Mauro (2018). "Digital constitutionalism: Fundamental rights and power limitation in the Internet eco-system", volume 80, issue 4, pp 295-301.

Pereira, Jane (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, ano 10, n. 35, p. 345-373.

Pernice, Ingolf (2015). "Global Constitutionalism and the Internet. Taking People Seriously", HIIG Discussion Paper Series, Discussion Paper Number 2015-01.

Pohle, Julia (2016). Multistakeholder governance process as production sites: enhanced cooperation "in the making". Internet Policy Review, vol. 5, 05-30.

Roberts, Sarah T. (2019). "Behind the Screen: Content Moderation in the Age of Social Media". New Haven, CT: Yale University Press.

Rue, Frank de La (2013). Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression. Human Rights Council, United Nations. Accessed March 6, 2020.

Scharpf, Fritz W. (1999). *Governing in Europe: Effective and Democratic?* Oxford: Oxford University Press.

Scharpf, Fritz W. (1996). Democratic Policy in Europe. *European Law Journal*, vol. 2, n. 2, 136-155.

Schmidt, Vivien A. (2013). Democracy and Legitimacy in the European Union Revisited: Input, Output and 'Throughput'. *Political Studies*, vol. 61, no. 1, 2-22.

Stivers, Camila (2018). "You Don't Know What You've Got Till It's Gone:" Rejoinder to Aaron Wachhaus's Article, 'Platform Governance'", Administrative Theory & Praxis, 40:150–158.

Suzor, Nicolas (2019). "Lawless". Oxford:Oxford University Press.

Suzor, Nick, Van Geelen, Tess and West, Sarah Myers West (2018). "Evaluating the legitimacy of platform governance: A review of research and a shared research agenda", The International Communication Gazette 2018, Vol. 80(4) 385–400.

Sylvain, Olivier (2010). "Internet Governance and Democratic Legitimacy", Federal Communications Law Journal, 62, 2, pp. 205-273.

Weber, Max (1919). "Politics as a Vocation", speech at Munich University (1918). Available at: http://fs2.american.edu/dfagel/www/Class%20Readings/Weber/PoliticsAsAVocation.pdf.
Wu, Tim (2010). Is Internet Exceptionalism Dead? In: Szoka, Berin & Marcus, Adam (eds): *The next digital decade: essays on the future of internet*. Washington: TechFreedom, 179-188.