On Legitimacy of Online Content Self-regulation: South Korea's News and Webtoon Case Study

ABSTRACT

To combat harmful content, online platforms have started to consider industry-level self-regulation. However, failed or half-buried self-regulatory attempts in TV, news, advertisement, and child online privacy in the U.S. have proved that coordinating self-interested stakeholders is fundamentally challenging. We conducted the first empirical research on online content self-regulation. Our focus is on South Korea, where self-regulation as a collaborative effort of creators and platforms is being actively pursued. We explored two popular industries, namely news and Webtoon. We recruited 15 participants with critical decision-making roles in these industries, including platform executives, creators, and government officials. By qualitatively analyzing stakeholders' attitudes and concerns, we figured that after self-regulation is initiated by a credible regulatory threat, self-regulators ought to keep gaining legitimacy through their expertise, procedural fairness, and effective enforcement mechanisms. The development of self-regulation is found to be highly contextual, depending on existing inter-stakeholder trust and the complexity of issues at stake. Based on the findings, we provided design principles for effective self-regulatory governance across content industries.

KEYWORDS

self-regulation, online content, online platforms, news, webtoon, regulatory governance, South Korea

1 INTRODUCTION

There is widespread concern that the online platforms failed to combat harmful content such as misinformation, hate speech, harassment, violent or sexually exploitative content [9, 13, 33, 36, 47]. Policymakers have tried to strengthen regulation on platforms, ranging from criminalizing harmful content [1, 8, 27] to increasing platform's liability [26]. However, this legislative approach is often criticized for compounding the chief virtue of the internet, freedom of expression [23, 25]. Accordingly, the United Nations [34], the White House [50], the European Union [12], and academics [3, 4, 14] call for an enhanced self-regulation of online platforms.

What is self-regulation? Self-regulation means rules imposed by the "regulatory target," which have been traditionally subject to state regulation [10, 14]. Conceptually, self-regulation can take place within or beyond an individual firm. In the context of online content, the first type of self-regulation is also referred to as "content moderation," which was heavily practiced and researched [17, 20, 22, 35].

But the second type, self-regulation through an organization representing industry stakeholders, has never been the center of attention. Recently, major platforms started to test out the second model by forming alliances to combat terrorist content [44] and protect minors online in EU countries [11]. Yet, no empirical research has been conducted on industry self-regulation as a tool of addressing harmful content. For brevity, this paper uses "self-regulation" only as a second meaning, unless otherwise indicated.

To fill this gap, we investigated South Korea, where online content self-regulation has been actively pursued for the past decades. We set up three research questions:

(1) How do industry settings affect the development of self-regulation?
(2) Why do industry stakeholders find self-regulation beneficial?
(3) What kind of self-regulatory governance would be perceived legitimate by industry stakeholders?

To answer these questions, We explored two popular industries, namely news and Webtoon, and recruited 15 participants with diverse and critical decision-making roles in these industries. Based on qualitatively analyzing in-depth interviews, we identified that self-regulation is heavily constrained by pre-existing industry setting such as cumulative trust between platforms and creators. Yet, we witnessed the overwhelming support of self-regulation as a defense mechanism against not only state's regulatory threat but

also demands of extreme users. Further, informed by the findings, we theorized basic design principles for legitimate self-regulation that serve the content industry and society at large.

2 BACKGROUND

History of failure. In the U.S., various media industries developed self-regulation, including TV [28], motion pictures [40], comic books [41], newspapers [5], and advertisements [38]. With very few exceptions, such as video game rating [15], these self-regulation models collapsed in the quickening period of the internet. Mainly, they could not receive cooperation and respect from stakeholders. Most self-regulators suffered from a lack of expertise and funds and could not effectively exercise their authority from the fear of losing members [28]. Stakeholders withdrew their support for or resisted self-regulatory decisions [41]. Furthermore, the government and courts accused self-regulation of anti-competitive conduct [2] and free speech violations [48]. Self-regulation did not survive this hostile environment.

Why South Korea? South Korea developed self-regulation across different types of content industries during its compressed growth from censorship under a military regime to an open internet under a consolidated democracy [24]. Armed with the support of courts and government agencies, Korea's content industries established a unique patchwork of participatory governance to tackle industry-specific problems like gambling (games), obscenity (films), and defamation (news) to preclude external regulatory attempts [32]. Creators, platforms, and civil society have been working together to create and enforce industry-wide rules.

Why news and Webtoon? Comparing two industries enables more enriched analysis with contextual factors. We selected news and Webtoon, major popular content that Korean people enjoy every day. Both industries faced serious regulatory threats in the past years. Concerns over misinformation triggered fake news law [7], and the government agency attempted to regulate violent Webtoon material [49]. Both attempts were frustrated by nationwide resistance from creators and rather gave rise to lively discussions about reconstructing self-regulation among industry stakeholders. We found these conditions fertile for an analysis of the triumphs and challenges of self-regulation.

3 METHODOLOGY

3.1 Theoretical framework

The survival of self-regulation depends upon stakeholders' willingness to support [45]. It is different from state regulation or company policy. Unpopular state regulation lasts until the law changes. A firm's investment in content moderation is mainly determined by a firm's leadership. Yet, stakeholders' participatory and autonomous decision-making, referred to as self-regulation's virtues, make it structurally vulnerable to the trust deficit of stakeholders.

We found two theoretical traditions especially useful to fathom the unique challenges of self-regulation. The first pillar is perceived legitimacy theory in law and political science. It offered us an "inside-out" perspective on the relationship between a self-regulator and stakeholders. Through this lens, we conceptualized self-regulation as the process of earning deference and cooperation from stakeholders by achieving certain factors such as the fairness of the procedure, and the consistency of outcome [21, 35, 45]. A second pillar is the rational choice theory in political economy that focuses on incentive systems of stakeholders in supporting rules [19, 46]. It enabled us to have a "outside-in" lens to understand why stakeholders would choose self-regulation over other alternatives.

Combining these approaches clarified that improving the legitimacy of self-regulation requires a multi-stakeholder analysis surrounding self-regulatory governance. With that, we decided to conduct in-depth qualitative interviews with creators, platform employees, self-regulators, and government officials to see how they evaluate self-regulation, what benefits and costs they are aware of, and what types of institutional design would be perceived as legitimate by them.

3.2 Interview protocol

3.2.1 Participants. We recruited individuals with critical decision-making roles in self-regulation in the news and Webtoon industries. As emailing official contacts did not work well, it took months to get in touch with those individuals. We reached out to creators and academics who appeared in the media about their work related to self-regulation and asked them to connect us to other entities. Oftentimes, people expressed their reluctance to share their experience due to potential repercussions, and we did not press further. Finally, we interviewed 4 creators, 4 online platform executives, 3 self-regulators, and 4 government officials, who either directly engaged in or closely observed the trajectory of self-regulation in their industries.

3.2.2 Procedure. 10/15 interviews were conducted in person in Seoul, South Korea, and 5/15 interviews were held remotely either by phone or by video conferencing program. The interview script was provided ahead of the meeting, and participants were allowed to delete or change questions. The first author of this work conducted in-depth interviews in Korean, ranging from 60 minutes to 100 minutes, and transcribed them. We obtained the informed consent of participants on the recording. Participants were notified that recordings would only be accessed by the interviewer. When

#Participant	Gender	Role	Industry	Affiliation
1	Male	Creator	News	-
2	Male	Creator	News	-
3	Male	Self-regulator/academic	News	-
4	Male	Self-regulator	News/webtoon (comments)	Korea Internet Self-regulation organization
5	-	Online platform	News	-
6	-	Online platform	News	-
7	Male	Government	News	Press Arbitration Council
8	Male	Government	News	Ministry of Culture, Sports, and Tourism
9	Male	Creator	Webtoon	freelancer
10	Male	Creator	Webtoon	freelancer
11	Female	Self-regulator/academic	Webtoon	Advisory Council on Webtoon Content
12	Female	Online platform	Webtoon	-
13	Male	Online platform	Webtoon	-
14	Male	Government	Webtoon	Korea Communications Standard Commission
15	Male	Government	Webtoon	Ministry of Culture, Sports, and Tourism

Table 1: Interview participants' demographic information. We used − to denote interviewees who were not comfortable with their information (e.g., gender, affiliation) being shared.

participants did not agree, the interviewer transcribed their answers during the interview.

We asked participants about (1) their views on online harm, (2) stakeholder relationships, (3) their evaluation of existing self-regulation and other alternative tools, and (4) solutions to improve the quality of self-regulation. We provided the complete interview script in Appendix A.

3.2.3 Ethics. We received an exemption from our institution's IRB, but we followed procedures for human research studies. All personally identified information was anonymized in the transcripts. Participants were compensated with a 30,000 KRW gift card (22 US dollars) exceeding the minimum hourly wage in South Korea (9,620 KRW). Some of them refused a gift due to their organization's policy. Participants were informed that they could stop or pause the interview at any time. After the interview, the complete transcripts were provided for their review.

3.3 Analysis and Translation

We chose a single-coder system instead of translating the entire transcripts to calculate the inter-rater reliability (IRR) for several reasons. The first is to avoid third parties access to the data, given the political sensitivity of interviews. The second is to minimize a loss in specific syntax and cultural nuances of an original language, following a recommendation for bilingual qualitative research [39]. Third, we believed that thorough discussions among authors on the formulation and revisions of the codebook would ensure the reliability and validity of the analysis.

Analysis followed two cycles of inductive coding and deductive coding. In the first cycle, the first author assigned Korean codes to the data using "concept coding" by labeling a macro level of meaning to data. Concept coding is known as useful for future category development in different languages [39]. After 3-4 transcripts first being coded, a set of

translated codes came up for discussion among authors. After the discussion, the coder moved on to the next segments of transcripts. Throughout this iterative process, authors had weekly meetings to refine and synthesize the codebook.

The second cycle of coding was conducted in English through a "theoretical coding" method to establish the hierarchy of concepts, as well as "evaluating coding," which symbolize participants' positive or negative attitude, to vividly capture participants' conflicting opinions on various concepts [37]. After finalizing the codebook, selective quotations were included after being translated and paraphrased. We provided the codebook in Appendix B.

3.4 Limitations

Similar to other qualitative studies, our findings are based on a small number of interviewees. In addition, due to the active state of self-regulation, we only recruited from South Korean news and Webtoon industries. Therefore, the results of our study cannot be generalized to the larger population. In addition, there is still a limitation of translated work. Although the first author is a native Korean speaker, the process of translation from Korean to English could have missed some nuances in the participants' responses. To improve the reliability of the findings, the authors discussed the themes and frequently referred to the original Korean transcripts to ensure that the translation captured the main insights in participants' responses.

4 RESULTS

We classified our findings into three categories: (1) Factual conditions in the industry that affected self-regulation, including stakeholder relationships and harmful content distribution; (2) Why stakeholders perceived self-regulation as a solution for combating harmful content; and (3) what organizational factors stakeholders would change to increase the legitimacy of self-regulation.

4.1 Pre-existing conditions that affect self-regulation

4.1.1 Many creators and centralized power of platforms (Q1). We observed the overwhelming agreement that the rise of the internet fundamentally changed both the production and consumption of online content. P1&2 (news reporters) and P10 (Webtoon creators) indicated the lower cost of entry, causing the explosive increase of creators. There are over 10,000 media companies and over 2,000 webtoon creators facing unprecedented competition for users' immediate attention. In addition, 7 participants pointed out the enormously centralized power possessed by online platforms. They stated that over 90% of users consume news articles and Webtoon through online platforms, and platforms make life-and-death decisions by highlighting or de-platforming content.

4.1.2 Webtoon creators trusted platforms, news reporters did not (Q2, Q5-6). While online platforms emerged as a custodian in both industries [22], the reactions of creators were different. According to P3 (news self-regulator), media companies perceived the rise of online platforms as their loss of direct channels with subscribers. P2 (news reporter) regrets the ever-increasing trend of selling sensational news titles to attract user clicks. P1 & 2 (news reporters) valued the professional ethics of the printing press and were worried about news consumption on online platforms that undermined investigative journalism. Likewise, P5 & 6 (online platforms) were aware of reporters' hostility against them but shared their frustrations of juggling conflicting demands from users and news companies.

> *P2 (News reporter): "If I could change one thing in the news history, I would choose the way [media companies'] are providing news to online platforms. We should have anticipated that news aggregation services would replace our direct communication channels with the audience."*

On the other hand, Webtoon creators perceived online platforms as partners, not adversaries. P 9 & 10 (Webtoon creators) generally appreciated online platforms' efforts to create new business models in the comics industry. Both creators considered themselves situated in a better position than their predecessors, who were poorly treated by comic book publishers. They also felt that their creativity was respected by platforms. Similarly, P 12 & 13 (Webtoon platforms) perceived creators as contributors and tried not to overstep their artistic freedom unless the illegality of content. P 12 offered artists annual medical check-ups, and P 13 provided free legal counseling services for creators for copyright disputes and other matters. In short, creators and platforms seemed to have created a reciprocal relationship.

4.1.3 Webtoon creators trusted other creators, but news reporters did not (Q2, Q5). Two industries also formed different levels of creator-creator trust. P2, a news reporter who has 30 years of experience, testified that he witnessed few successful collaborations between media companies because they considered others as competitors. P3, a self-regulator and former news reporter, said that journalists would suspect self-regulatory decisions in which other journalists engaged because they assumed other journalists prioritized companies' interests over their consciences. P2 prescribed that journalists' strong egos constrained their own ability to take collective actions to improve the quality of journalism. Thus, P1 (news reporter), P7 & 8 (government officials) were skeptical about the success of news self-regulation.

On the contrary, P 9 & 10 (Webtoon creators) expected other creators to represent overall creators' positions and did not suspect the possibility of them making self-interested decisions. In response to the inquiry about this difference in peer perception, participants estimated two reasons: (1) professional privilege: Comic artists had never been in a privileged position in society different from journalists; (2) the nature of market competition: while all news reporters pursue a single goal–breaking news, Webtoon creators do not directly compete with others due to the diversity of sub-genres and audience groups.

4.1.4 Identifying harm is challenging and subjective, with ambiguities and disagreements (Q1). Drawing boundaries of acceptable content is inherently difficult, but we found that the types of difficulties vary across industries. For example, P9 (Adult Webtoon artist) shared a question he faced, "is grown-up couples' school uniform costume play considered child pornography?" P11 (Webtoon self-regulator) received an increasing number of complaints about hate speech, such as gender discrimination or homophobic expression, which a traditional rating system could not resolve.

Meanwhile, P2 (news reporter), P3-4 (self-regulators), and P8 (government official) emphasized the difficulties of separating false or deceptive news articles: whether a report accurately captured fact, a reporter knowingly made a false report, or a reporter misled readers by covering monetary sponsorship. Fact-checking and making the judgment for these factual matters require laborious investigation and are likely to arise political debates. This characteristic of problems imposes more burden on news self-regulation compared to that of Webtoon, where adjudicators can make decisions on visualized material without long investigation.

> *P11 (Webtoon self-regulator): "We agreed in most cases, pretty quickly. It's sort of "you gotta know it if you see it.""*

In sum, stakeholders' attitude toward self-regulation depends on the context of the industry setting, including the

types of content being considered, their experience of stigmatization, and accumulated relationships among stakeholders. As we will see later, the insufficient trust among stakeholders impeded a self-regulator's ability to recruit its members and gain legitimacy for their decisions.

4.2 Perceived benefits of self-regulation

Self-regulation is expensive. P3 (news self-regulator) assessed at least 1 million dollars needed for news self-regulation. In addition to monetary costs, stakeholders may need to give up some authority to achieve industry-level coordination [28]. Nonetheless, our interview data demonstrate a strong agreement among all participants about the necessity of self-regulation. 6/15 participants who were skeptical about the prior performance of the existing self-regulator did not deny the necessity of self-regulation but argued for improving institutional arrangements. We were curious about why self-regulation draws overwhelming support among stakeholders. Like prior work [6, 14, 28], We expected that defending the industry against regulatory threats would be the primary reason, but participants offered more interesting explanations about multi-faceted benefits of self-regulation.

4.2.1 Individual stakeholder is vulnerable to users (Q3, Q5-6). We were struck by that 3/4 of creators mentioned "extreme users" as the most frightening source of external threat. P10 (Webtoon creator) shared his experience of stopping using "wa-geu-ra-no," which means "what happened?" in a dialect of the South East of Korea, because it was labeled as an insulting expression of the former President Roh Moo-Hyun.

> *P10 (webtoon creator): "We fear for users' censorship. As opposed to the government or a platform, user groups do not impose universal criteria. As internet communities became seriously polarized, communities have a very low tolerance for counter-views. Creators have faced lawsuits and personal attacks for their usage of 'prohibited words' defined by a certain group of users."*

P1 (news reporter) mentioned that he encountered a million-dollar lawsuit for his investigative report. The case was dismissed, but the procedure was burdensome to him and his advisors.

> *P1 (news reporter): "The purpose of government censorship in older days was well-behaved journalism. Yes, it is problematic, but at least they did not intend to end journalism. However, an aggressive and extreme group of contemporary users go after an individual journalist to death. It is a tangible, real threat."*

Users did not only threaten individual creators but also complained to platforms, self-regulators, and the government. P13 (online platform) and P11 (self-regulator) received an increasing number of complaints accusing certain expressions of hate speech or discrimination. In response to this concern, P5 (online platform) and P3 & P11 (self-regulators) conducted research about discriminatory or hateful expressions to set up rules to resolve increasing complaints.

Most participants were concerned about the chilling effect of the extensive list of words/symbols blacklisted by user communities. Traditional approaches such as rating and age verification are not well-equipped for this novel conflict, so participants largely agreed that self-regulation should provide a "safe space" to discuss these politically sensitive issues and help individual stakeholders to smoothly resolve user complaints.

4.2.2 Individual stakeholder is vulnerable to state's threats (Q2, Q3, Q6). All 15 participants agreed that self-regulation began in response to a credible threat from the state. The following is the list of interviewees' personal experiences related to the state's threats.

(1) P2 (news reporter) led labor movements to protect the freedom of the press in the 1990s.
(2) P1-2 (news reporters) opposed a Congress' proposal to enact punitive damages on defamatory news.
(3) P5 (online platform) and P4 (self-regulator) experienced individual politicians' unofficial requests to remove content.
(4) P4 (online platform) opposed a government proposal to regulate personal chats to reduce sexual crimes.
(5) P10 (webtoon creator) protested against the executive branch's publication ban on adult comics.
(6) P9 (webtoon creator) participated in a class action against the former President who put liberal artists on the blacklist of public funds.

Most surprising to us was that each participant conceptualized the state's threats in different ways. Prior research [14, 28] regarded perceived state threats as the government's regulatory actions, and P1-2 (news reporters), and P3 & 11 (self-regulators), P14 (government official) mentioned those cases. However, P5 & 6 (online platforms) and P3 (self-regulator) said that self-regulation was initiated by politicians' unofficial and groundless requests to remove content damaging their reputations. Political threats and repercussions led platforms to make an industry-wide collaboration to create an independent body to take political cases.

4.2.3 Collective peer reviews and conflict resolution benefit all stakeholders (Q3). Other than defensive functions, participants shared with us what they expected from self-regulation.

We extracted 73 quotations and classified them into five categories. A number in parentheses indicates the number of participants who mentioned the category.

(1) **Broad perspectives (11)**. Self-regulators make high-level rules beyond the perspective of an individual entity. For instance, P6 (online platform) appraised a self-regulator's rule-making to constrain public figures' removal requests.

(2) **Case accumulation (10)**. Self-regulators create a database for stakeholders' future reference based on industry-wide data on user complaints and regulatory decisions.

(3) **Conflict resolution (9)**. Self-regulators offer a forum where stakeholders can discuss issues at stake and find a mutual solution, which helps address inter-stakeholder disputes and deal with user complaints.

(4) **Peer review (6)**. Self-regulators foster peer review among creators. Creators would be less offended if peers tried to promote ethical standards.

(5) **Cost reduction (7)**. Self-regulators develop harmful content databases and harm detection technology that can be applied to multiple platforms.

We did not observe a meaningful difference between the news and Webtoon industries but did find a difference between creators and platforms. Creators valued self-regulation's function of "Peer review," while platforms appreciated "Broad perspective" and "Cost reduction." This result enabled us to conjecture an ideal look of self-regulation for which stakeholders would be willing to carve out their resources. The organization's cumulative decisions offer detailed ethical guidance and trigger a meaningful change of practice. This system would smoothly resolve disputes between creators, platforms, and users and enable big and small platforms to access advanced databases and technical tools. The remaining, more significant question is how to achieve it.

4.3 Design Principles for legitimate self-regulation (Q4, Q7-8)

Stakeholder relationships or user communities' behavior is not something that can be changed shortly. Thus, we tried to extract organizational factors of self-regulatory governance that policymakers can design in order to make self-regulation closer to an ideal look. We classified 320 statements into 7 categories and 20 subcategories, as demonstrated in Table 2. We clarified several categories that stakeholders disagreed upon. They include important factors such as who should make decisions, who should fund the costs, and how to police rule violations. We also found that these disagreements stemmed from the industry's pre-existing conditions.

4.3.1 Self-regulators must have a genuine attitude to serve for industry's common good. Most participants agreed that

Category	Code
Attitude	shared values
	room for creativity
Participation	creators
	platforms*
	neutral experts
	influential actors
	global actors
Expertise	individual expertise (law, media)
	collective rationality
Funding	creators
	platforms*
	government*
Due process	deliberation
	transparency*
	appellate processes
Incentives	carrots
	sticks*
Government Relations	respect for independence
	validation
	oversight*

Table 2: Design principles for legitimate self-regulation. We used * to indicate the categories participants disagreed upon.

self-regulators must be believed to pursue the same values as those of industry stakeholders. Especially, P9 & 10 (Webtoon creators) emphasized that a self-regulator must allow room for creativity instead of making rigid rules to suffocate freedom of expression.

4.3.2 Stakeholders' broad participation is important, but platforms joining self-regulation is contestable. No participant objected that: (1) creators and neutral experts should join a self-regulatory organization; (2) influential actors in the industry must take the lead of structuring it; and (3) global actors active in the Korean market should be included in it. Regarding global actors, P3 (self-regulator), P6 (platform), and P15 (government) mentioned their concerns about global online platforms' reluctance to contribute to "local" self-regulatory governance.

While P 9-10 (Webtoon creators) found online platforms' inputs helpful for protecting both platforms and creators, P2 (news reporter) and P3 (news self-regulator) disagreed with the idea of giving platforms a decision- making role. They thought that platforms could help enforce self-regulatory rules but not for creating rules or adjudicate cases. P3 suggested the idea that creators and platforms could establish "separate" self-regulatory organizations. This result reflects a platform-creator trust deficit in the news industry (See Section 4.1.2).

4.3.3 Self-regulators must have expertise in related content. Most participants in both industries stated that having reputable and qualified experts is critical to the legitimacy of decisions. Four participants, mainly from the news industry emphasized (1) each member's qualifications (attorney's license, journalism career) and (2) impartiality. 5 participants from both industries highlighted a decision-making structure that enable diverse experts to produce reasonable outputs which we called "collective rationality."

4.3.4 Stable funding is paramount, but funding sources are contestable. In our interviews, the minimum budget of news self-regulation (1 million dollars) were ten times as much as Webtoon self-regulation (100 thousand dollars). This difference seemed to be related to the amount of published content and the costs of monitoring and enforcement (See Section 4.1.4.

P2 (news reporter) emphasized that creators' share of expenses contributes to the independence of decision-making as well as the accountability of participating creators. However, due to creators' financial constraints, P2 stated that creators cannot be the only funding source. All 15 participants acknowledged that funding must come from either platforms or taxpayers. 3/4 online platforms, all 3 self-regulators, and 1/2 Webtoon creator supported platforms' contribution, concerned that government funding would entail political pressure. Meanwhile, all 2 news reporters and 1/4 government official disagreed with it because self-regulators will become biased toward platforms' business interests. Especially, both news reporters considered taxpayers' money safer than platforms' funding.

4.3.5 Due process is significant, but transparency is debatable. 12/15 participants agreed on the importance of deliberation. P2 (news reporter) participants warned that meritocracy of self-regulators often sacrificed the quality of decisions. However, there was a disagreement on transparency. 5 participants stated that transparency is a main vehicle of increasing the legitimacy of decisions. On the contrary, P3 & 11 (self-regulators) and P5 (online platform) were cautious about disclosing minutes of meetings because it would expose individual members to political attacks or lobby and impede conversations with candor among members. They were also worried about a rigid transparency rule impeding recruiting experts.

4.3.6 Self-regulation must have a credible and tangible incentive system. Incentives mean that reward for compliance and punishment for violation. As complying with standards is not always good for profits, almost all participants found it naive to expect all stakeholders voluntarily follow standards.

> *P2 (news reporter): "You need to be extremely realistic when it comes to structuring self-regulation. It*

is like steering wild horses. Journalists with strong self-esteem will never follow your lead unless you have very good carrots or sticks."

As a carrot, participants shared the ideas of publishing a credibility score or "safety" labels and giving more opportunities in government-funded projects and visiting scholar fellowships. As a stick, participants listed the disclosure of violations, monetary penalty, sharing the results of the investigation with government authorities, and more importantly, the exclusion of violating creators from online platforms. P2 (news reporter) and P3 (self-regulator), who were skeptical about the platforms' participation in self-regulation, cautiously supported the platforms' cooperation with enforcement because de-platforming media companies became the most effective punishment in the online news industry.

Meanwhile, P11 (Webtoon self-regulator) did not welcome having a more coercive authority. She was concerned that strict authority would compound the accumulated trust with industry stakeholders.

> *P11 (self-regulator): "I am skeptical about us having more powers against platforms because it would change our current partnership based on trust into unilateral regulation."*

4.3.7 The government should express their respect for self-regulation. Participants generally agreed that the government's official endorsement of self-regulation would increase the perceived legitimacy. In consistence with the prior research [14], all 3 self-regulators and 1 government official said that government oversight would motivate both self-regulators and industry stakeholders to support self-regulation.

> *P3 (self-regulator): "We stay motivated because the government would be willing to replace us anytime if we failed to perform our missions."*

However, 2 creators, 3 platform executives, and 1 government official emphasized that the government must keep a certain distance from self-regulation and must not exercise its authority unless self-regulation unquestionably fails.

5 DISCUSSION AND CONCLUSION

Civil society urges to establish more democratic internet governance [3, 43]. While "restoring" state's authorities over larger online platforms [30] is likely to overstep the virtue of free speech [3, 8], some scholars and practitioners suggest industry-wide self-regulation as an alternative [12, 14, 16]. However, no empirical research has been conducted, partially because similar attempts largely failed before the 1990s [6].

To fill this gap, we conducted qualitative research about South Korea's news and Webtoon self-regulation which has been actively pursued by industry stakeholders. Guided by

the concept of perceived legitimacy and rational choice theory, we took a multi-stakeholder approach to understand the key ingredients to self-regulation that make it work. We interviewed content creators, platform executives, government officials, and self-regulators in both industries.

In prior work, self-regulation was portrayed as a defense mechanism against the government's regulatory threat [14, 28]. It was true, but our interview data provided a much more enriched view of the self-regulatory ecosystem. As rational choice theory taught, the survival and success of self-regulation cannot be comprehended without analyzing benefits and costs that each stakeholder bear. A self-regulator ought to pursue finding a sweet spot where all influential stakeholders enjoy net benefits. This reciprocal relationship between stakeholders and a self-regulator create incredibly complex dynamics in self-regulation. Appendix C provides visuals of traditional and new frameworks.

Informed by the findings, we extracted pre-existing industry settings constraining and shaping self-regulation, the benefits of self-regulation from each stakeholder's perspective, and design principles for legitimate self-regulation. The main findings are as follows:

- The state regulation is not the only threat against free speech. Platforms suffer from politicians' informal content removal requests. Platforms, creators, government agencies, and self-regulators are stumped by extremist users' complaints. Thus, all of them need a "safe space" to discuss politically sensitive cases and alleviate their burdens of handling headaches.
- The cumulative trust between platforms and creators constrains self-regulatory system. When creators perceive platforms as adversaries, they resist against creating collaborative governance with platforms. Rather, they pursue creator-only self-regulation and seek government funding. However, due to a platform's "gatekeeper" position, creators would inevitably ask for platform's assistance in monitoring and enforcement. In contrast, when platforms and creators trust each other, they create collaborative governance, keeping a distance from the government.
- The difficulty of identifying harm affects self-regulation. For example, examining fake news and misinformation requires laborious investigation and entails political debates. Meanwhile, violent or explicit sexual depiction can be quickly determined. The former content needs a well-staffed self-regulator that prioritizes deliberation to efficiency. This finding is consistent with prior work that pornography was distinctively well-moderated among other harm on Reddit [36].
- A self-regulator ought to create an effective incentive system to derive industry-wide compliance. It is more

important to the industry with lower inter-stakeholder trust. Under the trust deficit, stakeholders are likely to be short-sighted and care less about their reputations. The gentleman's agreement would not work. The founders of self-regulator must design tangible carrots and sticks with the mindset of "steering wild horses."
- The financial reliance on either the government or platforms does not always endanger the independence of self-regulation. Prior work was concerned that government funding may open the door of political influence, while platforms' sponsorship may cause a conflict of interest [6]. However, both Webtoon (platforms' funding) and news (government funding) self-regulators were perceived as independent from funding sources. Neither funding sources are fundamentally right or wrong. It is more about cultural practices of respecting autonomy and expertise of self-regulators.

Our analysis demonstrates how communities thrive to resolve online harm based on participatory governance. After the impossibility theorem challenged technical methods of balancing competing values, technologists started to conceptualize value-oriented matters as a political question and resolve them through political approaches [29, 31]. Regarding online content, researchers explored creating platform governance based on democratic decision-making procedure [42, 51] and having an expert panel or community juries for content moderation [18, 35, 36]. Our case study moves further, showing that platforms became a co-founder of the national-level regulatory structure. We believe our findings have the potential to inform how computer scientists create next-generation trust and safety systems that might, by design, support the needs of industry-wide self-regulation.

While the computer science community has participated in the innovation of new technologies for creating and sharing digital content, those technologies do not exist in isolation. Rather, those technologies and content platforms exist in the context of content creators, content consumers, and society at large. A key component of "society at large" is the regulatory environment in which content sharing takes place. As both the computer science community and law and public policy community have vast experience studying technological and regulatory solutions, we believe that greater advances will happen when dialog happens across communities. By analyzing South Korea's online content self-governance, we view our work as contributing to that cross-community dialog.

REFERENCES

[1] 18 U.S.C. §§ 2257. 2006.
[2] United States v. NAB 536 F. Supp. 149 (D.D.C.). 1982.

[3] Jack M Balkin. 2017. Digital speech and democratic culture: A theory of freedom of expression for the information society. In *Law and Society Approaches to Cyberspace*. Routledge, 325–382.

[4] Hannah Bloch-Wehba. 2019. Global platform governance: private power in the shadow of the state. *SMU L. Rev.* 72 (2019), 27.

[5] Patrick Brogan. 1985. *Spiked: The short life and death of the National News Council.* Twentieth Century Foundation.

[6] Angela J Campbell. 1998. Self-regulation and the media. *Federal Communication Law Journal* 51 (1998), 711.

[7] Sang-Hun Choi. 2021. South Korea Shelves 'Fake News' Bill Amid International Outcry. https://www.nytimes.com/2021/10/01/world/asia/south-korea-fake-news-law.html.

[8] Danielle Keats Citron and Mary Anne Franks. 2014. Criminalizing revenge porn. *Wake Forest L. Rev.* 49 (2014), 345.

[9] Danielle Keats Citron and Helen Norton. 2011. Intermediaries and Hate Speech: Fostering Digital Citizenship for Our Information Age. *Boston University Law Review* 91 (2011), 1435.

[10] Cary Coglianese and Evan Mendelson. 2010. Meta-Regulation and Self-Regulation. In *The Oxford Handbook of Regulation*. Oxford University Press.

[11] European Commission. 2022. Alliance to better protect minors online. https://digital-strategy.ec.europa.eu/en/policies/protect-minors-online.

[12] European Commission. 2022. Self-regulation for a better Internet for kids. https://digital-strategy.ec.europa.eu/en/policies/self-regulation-better-internet-kids.

[13] Matthew Costello, Salvatore J. Restifo, and James Hawdon. 2021. Viewing anti-immigrant hate online: An application of routine activity and Social Structure-Social Learning Theory. *Computers in Human Behavior* 124 (2021), 106927. https://doi.org/10.1016/j.chb.2021.106927

[14] Michael A Cusumano, Annabelle Gawer, and David B Yoffie. 2021. Can self-regulation save digital platforms? *Industrial and Corporate Change* 30, 5 (2021), 1259–1285.

[15] Leyla Dogruel and Sven Joeckel. 2013. Video game rating systems in the US and Europe: Comparing their outcomes. *International Communication Gazette* 75, 7 (2013), 672–692.

[16] Evelyn Douek. 2020. The rise of content cartels. *Knight First Amendment Institute at Columbia* (2020).

[17] Evelyn Douek. 2022. Second Wave Content Moderation Institutional Design: From Rights To Regulatory Thinking. *Available at SSRN 4005326* (2022).

[18] Jenny Fan and Amy X Zhang. 2020. Digital juries: A civics-oriented approach to platform governance. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–14.

[19] Richard C Feiock. 2007. Rational choice and regional governance. *Journal of urban affairs* 29, 1 (2007), 47–63.

[20] Marta Cantero Gamito. 2016. Regulation. com. Self-regulation and contract governance in the platform economy: a research agenda. *European Journal of Legal Studies* 9 (2016), 53.

[21] James L. Gibson. 2012. Legitimacy Is for Losers: The Interconnections of Institutional Legitimacy, Performance Evaluations, and the Symbols of Judicial Authority. In *Motivating Cooperation and Compliance with Authority: The Role of Institutional Trust*. New York: Springer, 81–116.

[22] Tarleton Gillespie. 2018. *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media.* Yale University Press.

[23] Daphne Keller. 2021. Amplification and its discontents: Why regulating the reach of online content is hard. *Journal of Free Speech Law* 1 (2021), 227–268.

[24] You-Seung Kim. 2005. *Freedom and regulation on the information superhighway: A study of the Internet content rating system in South Korea.* University of London, University College London (United Kingdom).

[25] Jeff Kosseff. 2010. Defending Section 230: The Value of Intermediary Immunity. *The Journal of Technology Law and Policy* 15 (2010), 123.

[26] Mary Graw Leary. 2018. The indecency and injustice of section 230 of the Communications Decency Act. *Harvard Journal of Public Policy* 41 (2018), 553.

[27] Rachel Lerman. 2021. Social media liability law is likely to be reviewed under Biden. https://www.washingtonpost.com/politics/2021/01/18/biden-section-230/.

[28] Bruce A. Linton. 1987. Self-Regulation in Broadcasting Revisited. *Journalism Quarterly* 64, 2-3 (1987), 483–490.

[29] Zachary Lipton, Julian McAuley, and Alexandra Chouldechova. 2018. Does mitigating ML's impact disparity require treatment disparity? *Advances in neural information processing systems* 31 (2018).

[30] Tambiama Madiega. 2020. Digital services act. *European Parliamentary Research Service, PE* (2020).

[31] Thomas Miconi. 2017. The impossibility of" fairness": a generalized impossibility result for decisions. *arXiv preprint arXiv:1707.01195* (2017).

[32] Siho Nam. 2011. The World Narrow Web: Internet Content Regulation in South Korea. In *Cultural Identity and New Communication Technologies: Political, Ethnic and Ideological Implications*. IGI Global, 320–336.

[33] Matti Näsi, Pekka Räsänen, Atte Oksanen, James Hawdon, Teo Keipi, and Emma Holkeri. 2014. Association between online harassment and exposure to harmful online content: A cross-national comparison between the United States and Finland. *Computers in Human Behavior* 41 (2014), 137–145.

[34] United Nations. 2019. Strategies and Plan of Action on Hate Speech. https://www.un.org/en/genocideprevention/documents/advising-and-mobilizing/Action_plan_on_hate_speech_EN.pdf.

[35] Christina A Pan, Sahil Yakhmi, Tara P Iyer, Evan Strasnick, Amy X Zhang, and Michael S Bernstein. 2022. Comparing the Perceived Legitimacy of Content Moderation Processes: Contractors, Algorithms, Expert Panels, and Digital Juries. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022), 1–31.

[36] Joon Sung Park, Joseph Seering, and Michael S Bernstein. 2022. Measuring the Prevalence of Anti-Social Behavior in Online Communities. *arXiv preprint arXiv:2208.13094* (2022).

[37] Michael Quinn Patton. 2015. *Qualitative Research and Evaluation Methods (4th ed.).* SAGE Publications.

[38] Herbert J Rotfeld, Avery M Abernethy, and Patrick R Parsons. 1990. Self-regulation and television advertising. *Journal of Advertising* 19, 4 (1990), 18–26.

[39] Johnny Saldaña. 2021. *The Coding Manual for Qualitative Researchers (4th ed.).* SAGE Publications.

[40] John Alan Sargent. 1963. *Self-regulation: The Motion Picture Production Code, 1930-1961.* University of Michigan.

[41] Kevin W Saunders. 1994. Media self-regulation of depictions of violence: A last opportunity. *Oklahoma Law Review* 47 (1994), 445.

[42] Nathan Schneider, Primavera De Filippi, Seth Frey, Joshua Z Tan, and Amy X Zhang. 2021. Modular politics: Toward a governance layer for online communities. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–26.

[43] Olivier Sylvain. 2010. Internet governance and democratic legitimacy. *Federal Commmunication Law Journal* 62 (2010), 205.

[44] Global Internet Forum to Counter Terrorism. 2022. https://gifct.org/.

[45] Tom R. Tyler. 2012. Legitimacy and Compliance of Self-regulation. In *Legitimacy and Compliance in Criminal Justice*. Oxford University Press.

[46] Viktor J. Vanberg. 2002. Rational Choice vs. Program-based Behavior: Alternative Theoretical Approaches and their Relevance for the Study of Institutions. *Rationality and Society* 14, 1 (2002), 7–54. https:

//doi.org/10.1177/1043463102014001002

[47] Joe Whittaker, Seán Looney, Alastair Reed, and Fabio Votta. 2021. Recommender systems and the amplification of extremist content. *Internet Policy Review* 10, 2 (2021), 1–29.

[48] Writers Guild of Am., W., Inc. v. American Brdcst. Cos., 609 F.2d 355 (9th Cir.). 1979.

[49] Sung-won Yoon. 2015. Watchdog hit for excessive digital censorship. http://www.koreatimes.co.kr/www/tech/2021/03/133_176155.html?fa.

[50] Cat Zakrzewski. 2022. White House announces tech company efforts to combat violent extremism. https://www.washingtonpost.com/technology/2022/09/15/white-house-tech-extremism/.

[51] Amy X Zhang, Grant Hugh, and Michael S Bernstein. 2020. PolicyKit: building governance in online communities. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology.* 365–378.

A INTERVIEW PROTOCOL

The interview script is available at:

https://anonymous.4open.science/r/WebConf23-SelfRegulation-StudyMaterial/InterviewProtocol.pdf.

B CODEBOOK

The codebook is available at:

https://anonymous.4open.science/r/WebConf23-SelfRegulation-StudyMaterial/Codebook.pdf.

C COMPARISON OF SELF-REGULATORY FRAMEWORKS

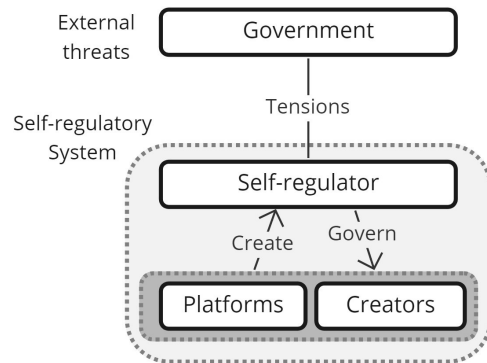**Figure 1: Traditional Framework of Self-regulatory System**

Figure 2: Multi-stakeholder Framework of Self-regulatory System