



Harmonizing the Cacophony with MIC: An Affordance-aware Framework for Platform Moderation

TANVI BAJPAI, University of Illinois at Urbana-Champaign, USA

DRSHIKA ASHER, University of Illinois at Urbana-Champaign, USA

ANWESA GOSWAMI, University of Illinois at Urbana-Champaign, USA

ESHWAR CHANDRASEKHARAN, University of Illinois at Urbana-Champaign, USA

Social platforms, and the online communities that use them, are evolving at a rapid pace. As a result, research and development regarding how to moderate online communities is being out-paced. In this paper, we present a novel framework that will allow moderation researchers and practitioners to not only keep-up with the diverse landscape of available platforms and affordances, but also comprehensively represent and analyze moderation on these platforms. The MIC framework represents a social platform's *moderation ecosystem* using a base-set of 12 platform-level affordances, along with a notion of the inter-affordance relationships that can exist between them. These affordances fall into the three categories—Members, Infrastructure, and Content—that are derived from Grimmelmann's taxonomy of moderation. We demonstrate the advantages of using an affordance-aware framework like MIC by analyzing several social platforms over the course of two case studies. First, we analyze individual platforms using MIC and demonstrate how MIC can be used to examine the effects of platform changes on the moderation ecosystem and identify potential new challenges in moderation. Next, we use MIC to systematically compare three platforms and propose potential moderation mechanisms that each can adapt. Moderation researchers and stakeholders can use such comparisons to uncover where platforms can emulate the moderation practices of successful, established, and better-studied platforms, as well as learn from the pitfalls other platforms have encountered.

CCS Concepts: • **Human-centered computing** → **Computer supported cooperative work**.

Additional Key Words and Phrases: social media, platform development, moderation ecosystem, online communities

ACM Reference Format:

Tanvi Bajpai, Drshika Asher, Anwesa Goswami, and Eshwar Chandrasekharan. 2022. Harmonizing the Cacophony with MIC: An Affordance-aware Framework for Platform Moderation. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 385 (November 2022), 22 pages. <https://doi.org/10.1145/3555110>

1 INTRODUCTION

The moderation of online communities has been the focus of a large body of social computing research [13, 17, 24–29, 39, 41–43, 53]. Much of this research is unified by the use of Grimmelmann's taxonomy of moderation [20], which defines, in generality, the terminology, goals, techniques, and challenges of moderating online communities. Grimmelmann characterizes an online community using three features: the community's *members*, the *content* that is shared among the members, and the *infrastructure* used to share it. Grimmelmann's four techniques for moderation, *excluding*, *pricing*, *organizing and norm-setting*, are all defined in a way that does not factor in the diversity in communities and the technologies they employ. For example, *exclusion* refers to keeping unwanted members from joining the community. On Reddit, this is done by banning problematic users from

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.

2573-0142/2022/11-ART385 \$15.00

<https://doi.org/10.1145/3555110>

a community (subreddit) page [23]. On Discord, community (server) administrators may require members to verify that they have a domain-specific email address; to do so, they can create custom authentication applications using Discord developer tools.¹

Grimmelmann's taxonomy is unequivocally useful for unifying moderation across communities and technologies. However, in its current formulation, Grimmelmann's taxonomy does not explicitly account for nuances at the platform-level (e.g., affordances, technological or design related). It is clear that the Social Networking Sites (SNSs), or social *platforms*, that online communities use shape many of the moderation challenges they face, as well as the strategies they employ to address them. In particular, it is the platform's *moderation ecosystem*, i.e., the components that impact moderation and the interactions among them, that play a central role in how the communities that use them are moderated. As such, recent moderation research is centered around particular platforms (e.g., [25, 41]), or newer technologies such as Virtual Reality (VR) [9]. As more platforms are created and updated, so too are the moderation strategies, needs, and challenges of the online communities that use them. For instance, communities that use voice rooms [25] or VR technology [9] face novel moderation challenges unlike those of communities on more traditional and better-studied platforms such as Reddit or Facebook. Hence, it is incredibly important to research and develop moderation on newer platforms as they emerge and become more popular.

Unfortunately, platform development typically outpaces moderation research. A clear example of this can be seen in the recent rise in popularity of audio-based social platforms such as Clubhouse, which launched into the mainstream during the global COVID-19 pandemic [44]. Clubhouse's success was closely followed by the introduction of other audio-focused platforms and extensions to existing platforms [38]. Sonar, an alternative voice-chatting app, launched in January 2021 [45]. Both Twitter and Facebook launched live audio room features during 2021. Other popular platforms such as Reddit [37], Telegram [54], Slack [40], and Discord [3] quickly followed suit and began launching their own Clubhouse-esque features. During this time, Spotify acquired the parent company of an audio-only, sports-centered app called Locker Room [4, 46], and later re-branded and re-launched it as direct competitor to Clubhouse called Spotify Greenroom [11]. A timeline of this audio-based platform "boom" can be found later on in Figure 1.

1.1 Moderation Research in the Landscape of Evolving Social Platforms

Similar to the development of any new social technology, questions about moderating such platforms continues to be of particular interest to not only the Computer-Supported Cooperative Work and Social Computing (CSCW) research community, but also other moderation stakeholders such as platform designers, moderators, and community members. We identify three key challenges that researchers and stakeholders face when addressing moderation in the landscape of dynamically evolving social platforms. First, it may be tempting to choose one or two representative platforms while developing new insights to their moderation when, in reality, these platforms are diverse in ways that affect moderation. For instance, despite their similarities, Spotify Greenroom allows users to enable a text-based chat box in their live audio room while Clubhouse does not. Secondly, many of the new platforms or features might appear to be novel or unstudied, when they are essentially repackaged versions of older and more established technologies. Spotify Greenroom's chat boxes are similar to those from Twitch, a popular live video-streaming platform; Sonar's world-building concept resembles classic virtual world building games such as Minecraft. And finally, as was previously discussed, these platforms are rapidly evolving and adding features that not only impact moderation but also could undermine or out-date moderation research by the time

¹An example of one such Discord tool can be found at <https://github.com/sigpwny/sigpwny-shibboleth-auth>

it gets published or implemented. For example, Clubhouse added several new features that impact moderation during the time between this manuscript's submission and publication.

To address these challenges, and better enable the moderation research community to keep up with rapid platform development, we develop a new theoretical framework to aid in the analysis of moderation on social platforms. Our framework can benefit moderation stakeholders by enabling them to identify potential moderation challenges they could face on the platform, as well as adapt or design moderation solutions to address them.

1.2 The MIC Framework

In this paper, we present a novel theoretical framework that allows us to represent social platforms' *moderation ecosystems*. This representation is comprised of a base set of twelve relevant platform-level *affordances*. Each affordance falls into one of three categories that are derived from Grimmelmann's [20] definition of an online community: **M**embers, **I**nfrastructure, and **C**ontent. As such, we call our framework MIC. MIC is also able to represent how these affordances could impact each other by defining a notion of *inter-affordance relationships*.

The MIC framework has key implications for moderation researchers and stakeholders. More concretely, we argue that the advantages of using the MIC framework are three-fold:

- (1) The affordances and inter-affordance relationships in MIC provide a simple and explicit representation of potentially complex or subtle moderation ecosystems of social platforms. These components will also provide moderation researchers and community owners a convenient "checklist" to aid them in exploring and considering platforms to understand how moderation occurs on them.
- (2) MIC can be used to compare and contrast platforms' moderation ecosystems. Online community owners can use these comparisons to help decide which platforms would be more conducive for the moderation needs of their communities. Moderation researchers and platform designers can use these comparisons to uncover where platforms can adapt and learn from more established and better-studied platforms, as well as learn from the pitfalls these platforms have encountered.
- (3) MIC's representation of a platform's moderation ecosystem can be easily updated to reflect platform changes. Inter-affordance relationships can also be examined to catch potential moderation issues that new features could cause. This will allow moderation researchers and stakeholders to update their understanding of platforms, and re-evaluate and potentially update moderation strategies and tools that might be impacted by platform changes.

We will use MIC to analyze several social platforms through two case studies to exhibit these advantages. Our first case study focuses on analyzing an individual platform using MIC, and shows how MIC can easily reflect platform changes as well as propagate such changes throughout the moderation ecosystem to account for new moderation challenges. In the second case study, we use MIC to systematically compare three platforms and use these MIC-based comparisons to propose potential moderation mechanisms that platforms can adapt from one another.

2 BACKGROUND

To better motivate our theoretical framework, we will first outline the relevant components of Grimmelmann's taxonomy [20], since our framework builds on top of these components. Then, we will briefly introduce each affordance while discussing the previous moderation work that informed our decision to include it in MIC.

2.1 Grimmelmann’s Taxonomy of Moderation

Grimmelmann defines an online community using three features: the community’s *members*, the *content* that is shared among the members, and the *infrastructure* used to share it [20]. We use these features to delineate the three main categories for affordances that are included in the MIC framework. We opt for this organization because each facet of this definition can impact moderators’ ability to carry out Grimmelmann’s four basic techniques for moderation [20]. *Exclusion* is the act of excluding problematic or unwanted members from the community. Another closely related technique is *pricing*, which controls the participation of community members by introducing barriers to entry. Both exclusion and pricing are mandated by the infrastructure and members of the community: infrastructure provides the tools for exclusion or pricing, while members are involved in using these tools. *Organizing* is a technique that involves “shaping the flow of content from authors.” This technique is not only impacted by the nature of content within the community, but also by the infrastructure that provides certain members with such “shaping” capabilities. Finally, *norm-setting* involves the creation and articulation of community norms to establish acceptable types of behavior. Norm-setting is often accomplished by the other techniques, and by identifying and highlighting members of the community that exhibit these types of behaviors. In the following subsections, we will describe findings in moderation literature that speak to these claims.

2.2 Member-related Affordances

Through interviews with volunteer moderators of Discord servers, Jiang et al. [25] found that server owners create custom user roles to distinguish between various *user types*. The moderator role is a common facet of online communities and a role that is often assumed by volunteers on platforms relying on distributed moderation [17, 25, 43, 55].

The second member-related component in our framework is *anonymity*. Schlesinger et al. [41] studied how anonymity affects content on Yik Yak, a social media application that allowed users to make anonymous text posts that are grouped by location [41]. In general, anonymity has been found to have both positive and negative effects on social interactions [15]. Outside the context of online social spaces, anonymity was found to remove status markers that prevent members from participating in discussions on collaborative systems [21, 33, 52]. Prior work examining the role anonymous voice-based interactions in online games found that in some cases anonymity was lost due to the nature of voice-based communication, and this caused some players to feel uncomfortable [51]. In fact, this loss of anonymity was deemed as one of the main reasons behind gamers abandoning the game being studied.

2.3 Infrastructure-related Affordances

One of the main infrastructural affordances we consider is a platform’s *organization*, i.e., how content and communities of the platform are situated. On Twitch, text-chats are associated to specific live streams, and live streams are separated by different Twitch channels; different streaming channels have different moderators and moderation strategies [42]. In certain cases, the lack of certain organizational structures within platforms might force communities to use a combination of two or more platforms to overcome these deficiencies. This might lead to various *inter-platform relationships*, which can be seen in prior work studying how moderators of Reddit communities use both Reddit and Discord to host their communities and the resulting challenges faced by multi-platform communities [26].

Other integral parts of infrastructure include the *rules and guidelines* of platforms and the communities they host. Prior work has examined the rules that moderators of both Reddit and Discord outline for their communities, as well as guidelines specified by the platform itself [25, 26].

Rules and guidelines, both community-defined and platform-specified, often describe the different roles members can play within the community (e.g., both Discord and Reddit have online manuals to help guide volunteer moderators. Platform and community rules and guidelines have also been shown to play a large role in shaping community norms [16, 18, 27, 49]. Platforms often have different *badges and markers* to help users identify quality content or experienced users, and can be used to shape the behavior of users [6]. A common challenge encountered in video-based or voice-based communication systems is a lack of markers that provide cues to indicate when a user wishes to speak [22, 35]).

We designate a *moderation mechanisms* affordance to represent a platforms' designated moderation tools or features, i.e. the infrastructure that a platform provides specifically for moderation. Reddit has automated moderation tools, as well as an API that allows moderators to create moderation tools and bots to help human moderators to review large volumes of content. Discord has similar tools for moderators, some of which have been found to cause unprecedented moderation issues [25]. Prior work has explored how volunteer moderators employ a variety of mechanisms for moderating content, and moderation typically involves a large amount of time and effort to keep up with the massive amounts of content generated within social platforms [26, 32]. As a result, automated and human-machine collaboration tools have been developed to assist moderators on text-based platforms like Reddit [12, 23]. Video-hosting platforms like YouTube use algorithmic moderation to allow for a larger moderation purview and to alleviate the labor of human moderators [19, 39]. Finally, platforms that enable *monetization* may have unique moderation problems, since monetization has been found to lead to controversial behavior online to achieve virality [8], and algorithmic moderation tools can negatively impact users who rely on monetization revenue [34].

2.4 Content-related Affordances

Our framework considers the various *modalities* platforms can support. As discussed in the previous subsections, the modality of content plays a role in how the content is viewed, organized, and moderated. Voice-based communication systems and audio-based communication used in online gaming utilize real-time, or *synchronous* audio [5, 47, 51]. Ackerman et al. [5] studied how users viewed and used Thunderwire, a collaborative audio-only real-time communication system modeled after telephone "party lines" of the late 19th century. Wadley et al. [51] studied real-time audio-communication in online multiplayer games and virtual worlds during game play. However, there are voice-based communities from India that use asynchronous audio for communication [36, 50].

An affordance closely related to synchronicity is *ephemerality*, as it is often, but not always, a consequence of synchronous or real-time content. Both communities studied by Ackerman et al. [5] and Wadley et al. [51] used synchronous and ephemeral content. Prior work on ephemerality in social platforms has largely focused on ephemerality of text posts, links or images [7, 41, 56]. Jiang et al. [25] studied the challenges of moderating voice on Discord and found that the synchronicity and ephemerality of audio-based content created novel challenges for moderators.

Finally, social platforms can allow for certain *access and restrictions* imposed on either viewing or creating content. In the past, subreddit moderators have purposely restricted access to their content as a way to express dissatisfaction with certain platform changes [31]. Restrictions and access have been used to subdue antisocial behavior, though the efficacy of doing so is largely unclear [48].

3 MIC: A FRAMEWORK FOR REPRESENTING THE MODERATION ECOSYSTEM OF SOCIAL PLATFORMS

In order to formally define MIC and provide examples of its affordances and inter-affordance relationships, we will use three audio-based social platforms (Discord, Spotify, and occasionally Soundcloud) as working examples. These three platforms pre-date the audio-based social platform

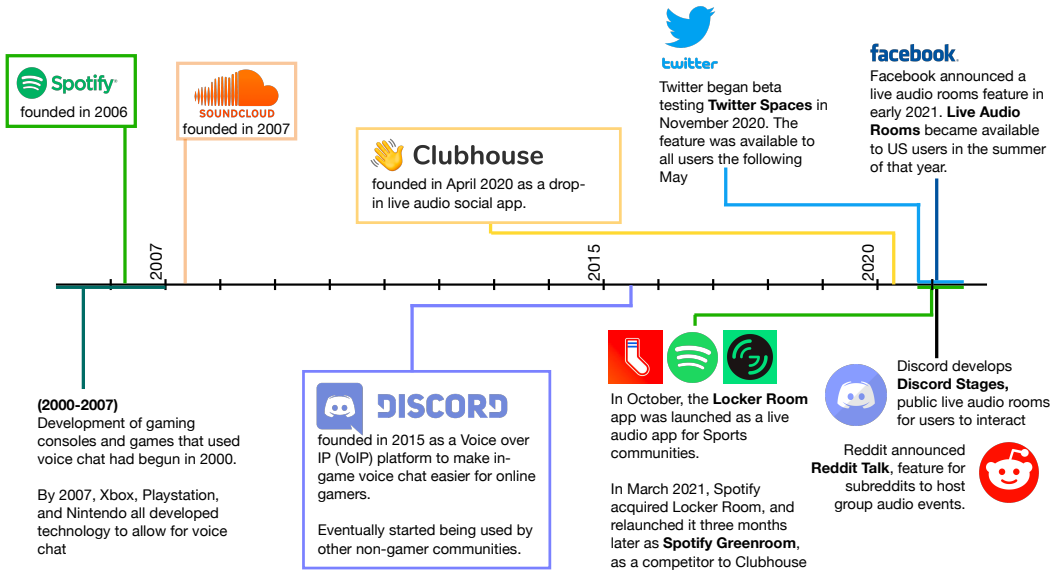


Fig. 1. A timeline of popular audio-based technologies and social platform development. Clubhouse appears to mark the beginning of an audio-based “boom” in platform development, however platforms have utilized audio content since as early as 2000. We focus our attention on audio-based social platforms Spotify, Discord, Clubhouse, and occasionally Soundcloud, to showcase MIC’s utility.

“boom,” which can be seen in the timeline shown in Figure 1. For Discord and Spotify, we will construct *MIC diagrams* (see Figures 2 and 3) to better highlight how MIC represents their moderation ecosystems and inter-affordance relationships that exist within them. These MIC diagrams, as well as one we will construct for Clubhouse in Section 4, will be used to guide platform comparisons in Section 5. High-level descriptions of Spotify, Discord, and SoundCloud are provided below.

Spotify. A audio-streaming service that hosts both music and podcasts. The main two types of Spotify users are listeners (those who use the service to stream content) and creators (those who use the service to upload content). Listeners are able to follow both creators and other listeners, and can view the latter’s playlists and listening history. Creators must use other Spotify services, such as Spotify For Artists (for musicians) and Anchor (for podcasters).

SoundCloud. A music-sharing website that allows all users to post audio (which consists of music, podcasts, random noises, etc). Users are able to comment on audio files and re-post others’ audio posts on to their feed.

Discord. A messaging platform that allow users to communicate via text, voice, or video. Discord’s infrastructure is composed of “servers,” which can be thought of as landing pages for individual communities that use the platform. Servers can contain topic specific text-channels or voice/video channels. Server owners can create custom roles for server members, and can associate specific permissions for each role.

3.1 MIC Affordances

For each of MIC’s twelve affordances, we provide a general description and highlight how they are supported by different platforms through our working examples. We will continue to discuss how these affordances play a role in moderation on platforms.

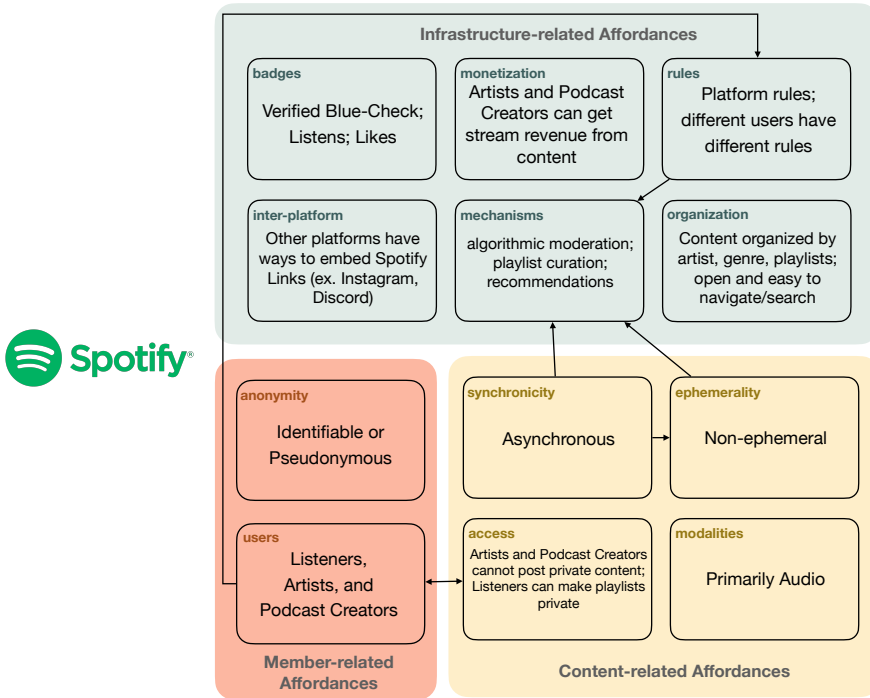


Fig. 2. The MIC diagram for Spotify provides a graphical representation of its moderation ecosystem, and shows the inter-affordance relationships that occur using directed arrows. Moderation on Spotify is done primarily by the platform itself (**mechanisms**), as there is no moderator role available to users (**users**), and content can only be removed by the poster or the platform (**access**).

Modalities (modalities). *Unimodal* platforms are centered around one primary type of modality, while *multimodal* platforms use multiple types of modalities. Discord is multimodal since servers contain text- and voice/video- channels. Spotify, on the other hand, is unimodal since audio is the primary type of content supported by the platform. The existence of multiple modalities will affect moderation on the platform, since having more than one modality typically requires a broader set of policies and tools for moderation [25, 26, 32].

Access and Restrictions (access). Platforms often have various access and permission settings that allow or prohibit content from being posted, viewed, or removed. Many of these settings are accessible by the content creator, while some are limited to the platform. Discord allows server-owners and moderators to limit access to the server itself and to channels; the ability to use certain messaging features can also be limited by owners or moderators. Spotify only allows creators (musicians or podcasters) to publish content. Since Anchor is a free service for users who wish to become podcasters, there are no restrictions to post podcasts. However, users cannot publish music to Spotify directly—they must use a music distributor. Popular musicians are often signed to record companies or labels that will either act as or employ a distributor. Independent artists, those who do not have the backing of a record company, can use online music distribution services² to publish music on Spotify. These services are never free, and therefore access to publishing music on Spotify is restricted. SoundCloud allows all of its users to post audio-content, and only limits the amount

²DistroKid (<https://distrokid.com/>) is one such distribution service.

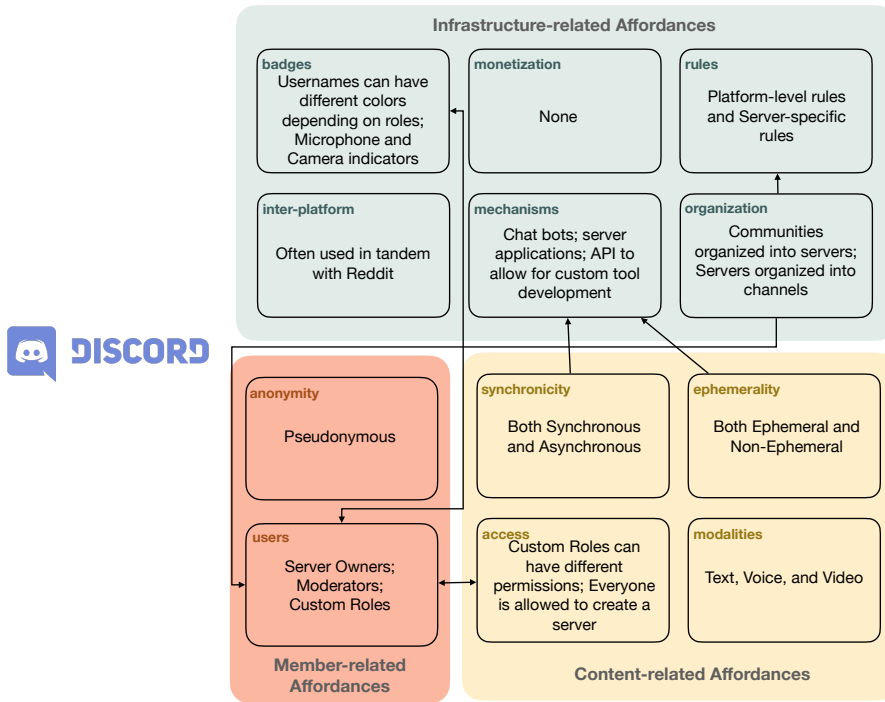


Fig. 3. The MIC diagram for Discord. Moderation on Discord is done primarily by users who assume the moderator role in each server (**users**); moderators can use tools provided by the platform itself, or create their own tools to address their community's needs (**mechanisms**).

of audio-content a free user can upload before requiring a paid SoundCloud Pro account. Spotify's and SoundCloud's access barriers are examples of Grimmelmann's pricing technique [20].

Monetization (monetization). Monetization refers to whether a platform provides infrastructure that enables users to generate revenue from their participation. Discord does not provide built-in mechanisms to enable monetization. Spotify monetizes all music, paying the artists according to how many times a song is streamed. Spotify also provides monetization tools through Anchor to allow podcast hosts to monetize their content as well. Content that is being monetized may be more heavily moderated than content that is not; monetization may also incentivize creators to generate more content, which could lead to more moderation challenges.

Synchronicity (synchronicity). A platform's content is synchronous if it is being generated in real-time. Voice chats on Discord can only occur synchronously, whereas text-based conversations may occur asynchronously. Audio on Spotify is asynchronous. Synchronous content often creates challenges for moderators since not all moderators or moderation mechanisms can be present at the time the content is being created. Asynchronous content provides a larger window of opportunity for moderation mechanisms to detect and report abusive content or behavior.

Ephemerality (ephemerality). Ephemerality refers to whether or not the content can be accessed after it has been created and/or posted. On Discord, voice chats are ephemeral, since recording voice-channels can violate Discord's Terms of Service. On Spotify, audio is not ephemeral. Studies have shown that users behave differently when interactions are ephemeral and leave no record

or trace [7, 41]. Furthermore, when content is ephemeral, it becomes difficult for moderators to collect robust evidence to prove that anti-social behavior occurred to remove bad actors [25].

User Types (users). Platforms may distinguish between types of users, and may even have designated types that allow users to act as moderators. Different user types are often associated with different permissions. On Discord, server owners and administrators can create custom roles for users, each with custom permission settings; one such role is typically assigned to “moderators.” On Spotify, only users with Spotify for Artist accounts are able to publish music. All users are able to create Anchor accounts to publish podcasts. Spotify has no designated “Moderator”-like role assigned to users on the platform.

Anonymity (anonymity). Users on platforms may be anonymous or use pseudonymous usernames to mask their identity. On Discord, users typically adopt usernames or handles that are custom and/or pseudonyms. Thus, users in voice-channels might not be not associated with any actual means of identification. On Spotify, listeners can, and often do, create account usernames with their actual identity (typically by linking Spotify to other social media accounts, such as Facebook). However, some users do adopt custom usernames that obscure their identity. Creators may publish audio-content under stage names or aliases. Anonymity has been found to both enable and discourage negative behavior in online social spaces [21], and anonymity appears to break down when using voice-based communication [51].

Organization (organization). The organization of a platform refers to the way in which content and communities are organized, situated, and discovered on the platform. A platform’s organization impacts users’ and moderators’ ability to locate content and members of interest. Discord is organized into servers, and each server has various channels in which community members interact and share content. Users can use Discord’s Server Discovery feature or Explore page to look for popular public servers to join, or create their own public or private servers. Not all large servers are necessarily public or searchable using the Server Discovery feature. The vast majority of audio-content on Spotify is indexed and publicly available to every user of the service. Typically, audio on Spotify is organized by artist, genre, podcast, or in user- or algorithmically-curated playlists (some of which are private). Users can search and discover all public audio-content via search or using Spotify’s various discovery and recommendation mechanisms.

Rules and Guidelines (rules). Most platforms utilize some combination platform-wide terms of service (TOS) and community-specific guidelines to govern user behavior. These terms and guidelines establish high-level rules that all users are expected to abide by. In addition to community guidelines and TOS, Discord also has platform-level rules that clearly define the roles of moderators on servers. At the community-level, Discord servers can publish their own set of rules and guidelines that are typically more tailored to the type of community the server hosts. Spotify has separate guidelines and TOS for listeners and content creators who use Spotify for Artists and Anchor. The rules and guidelines help establish a baseline for both platform-wide and community-specific norms and conditions for exclusion (e.g., suspensions or bans [13]). Rules and guidelines play a key role in moderation, as seen in Grimmelmann’s work—*norm-setting* and *exclusion* make up two of the four common techniques for moderation [20].

Badges and Markers (badges). Badges and markers refer to the various types of visual cues or indicators that could be applied to users and content. On Discord, different user types can have different colors associated with them. For example, if a “moderator” role is associated with the color red on a Discord server, we know that a user’s handle (i.e., username) appearing in red indicates that the user is a moderator. Such markers help other members identify the official moderators

of a server, and depending on what other roles the server defines, could help identify different types of users. Discord also provides indicators that show whether participants of a voice call have their microphone muted or their video on; this information can be seen without having to actually join the voice-call. On Spotify, a “verified” blue-check mark on an artist’s profile indicates that the identity of the owner of the artist page has been officially verified by Spotify. This signal indicates to users that the content posted on this artist’s page is coming from an official source. Spotify also displays the number of times a song has been listened to and the number of users who have liked a playlist. Such badges and markers help in moderation since they provide users and moderators with additional cues to determine whether certain users or content are safe to engage with.

Inter-Platform Relationships (inter-platform). The way users of one social platform utilize other platforms is an aspect that is not often highlighted when discussing moderation on social platforms in general. Discord servers are known to be used alongside other platforms (such as Reddit [26]), but are also commonly used alone. Discord users will occasionally use other, more free-range platforms such as Twitter and Reddit to discover and advertise private servers. Spotify, on the other hand, is often used by other platforms to embed music. For instance, Instagram users can add music directly from Spotify to their story posts, or link to their Spotify playlists. As more SNSs become available, it will be more commonplace for online communities to use more than one platform. This affects moderation since bad actors can harass users across multiple platforms, making moderation more difficult [24].

Moderation Mechanisms (mechanisms). The moderation mechanisms of a platform refer to its built-in moderation tools and procedures. Discord allows users to use and create chat bots and tools to moderate text-channels. Discord also has a guide for moderators. However, not all interactions in a voice-channel can be moderated unless a moderator is present in the voice-channel every time there is activity or the voice-channels are being recorded. Discord has bots that enable recording, but depending on where users reside, consent must be granted in order for recording to be allowed. On Spotify, all audio content *can* be moderated by the platform itself, since audio must be first uploaded to the platform and processed before it is hosted publicly. Spotify has mechanisms for algorithmic content moderation; this is the case with moderating copyright-abiding content [10], and the existence of such mechanisms leads us to believe that all audio-content is moderated in some way. Limited moderation mechanisms allow abusive and antisocial behavior to go unchecked on social platforms.

3.2 Relationships Between Affordances

Though we have defined a set of disjoint affordances, affordances can have relationships with each other. For instance, in both Spotify and Discord, access is linked to user roles, since different types of roles constitute different types of access. Highlighting these inter-affordance relationships will illuminate how potential modifications to one affordance could impact the moderation ecosystem at large. Moreover, if a specific affordance has been identified as a contributor to moderation challenges, we can use inter-affordance relationships to identify other, less apparent affordances that also contribute to these challenges.

Formally, we define an inter-affordance relationship from affordance *A* to affordance *B* if modifying affordance *A* impacts or changes the status of affordance *B*. For example, the asynchronous nature of content on Spotify (**synchronicity**) enables its non-ephemerality (**ephemerality**); indeed, if Spotify introduced synchronous content, then the ephemerality of certain content might change³.

³In fact, Spotify Greenroom has synchronous and ephemeral content, and gives users to option to record live audio rooms to upload to Spotify.

On Discord, the **ephemerality** and **synchronicity** of the voice interactions in voice-channels affect the moderation mechanisms that are available on the platform. In our MIC diagrams, these relationships are shown as directed arrows between affordances. A bi-directional arrow is used to indicate when a relationship exists in both “directions.” For example, **user** types on both Spotify and Discord are tied to types of **access** and permissions. These relationships in a platform will likely change over time as the platform itself is updated.

Other inter-affordance relationships in Spotify and Discord’s moderation ecosystems are as follows: The non-ephemeral (**ephemerality**) and asynchronous (**synchronicity**) nature of content on Spotify affects the platforms’ moderation **mechanisms**. Similarly, the moderation **mechanisms** are enabled by Spotify’s user agreement, which explicitly states that the platform is allowed to remove or edit any content that is uploaded if it violates community guidelines (**rules**). On Discord, user types are often customized to each server, thus the **organization** of Discord has an affect on **user** types.

4 MIC AS A TOOL FOR ANALYZING INDIVIDUAL PLATFORMS

In this section, we will demonstrate how MIC can be used to represent and subsequently update our understanding of a particular platform’s moderation ecosystem. We will use MIC to analyze the rapidly evolving Clubhouse app at two different points in time.

First, we will describe Clubhouse and its affordances as of June of 2021, and in doing so will construct its MIC diagram (Figure 4). Then, we describe the changes made to Clubhouse between June of 2021 and January of 2022 that could potentially impact moderation, updating the MIC diagram accordingly (Figure 5). Finally, we will discuss how using MIC allows us to reason about moderation strategies and challenges that exist on Clubhouse in a more efficient and systematic way, and what insights MIC provides that may otherwise be overlooked.

4.1 Exploring Clubhouse Using MIC

Clubhouse (as of June 2021). Clubhouse is an invite-only, so new users had to be invited to the app using their phone number (**access**). The platform’s community guidelines require users to use their real names (**anonymity**). Clubhouse users can only communicate with one another using audio in public or private voice rooms (**modalities**). Clubhouse is organized into topic-specific pages and groups called “clubs” (**organization**); only “the most active members of the Clubhouse Community” can create clubs (**access**). Each such page and club is made up of synchronous and ephemeral voice rooms (**synchronicity, ephemerality**). Every club has designated admins that have the ability to edit the club settings, name, and manage members (**users**). Public voice rooms can be accessed by any user on the app, regardless of their membership in its associated club or interest in the room’s subject (**access**). Private rooms can only be joined by the followers of the room host or the members of the room’s associated club (if it exists) (**access**). All participants of rooms are required to follow Clubhouse’s Community Guidelines [2] (**rules**). However, established clubs can publish a list of club-specific rules that can be applied to participants of rooms hosted by the club (**rules**).

Users can have one of three roles in a room on Clubhouse (**users**). The moderator role (denoted by a green star symbol) is given to the user who creates the room. This user has the ability to end the room, invite users to the stage to speak, mute speakers, and assign other users to be moderators as well. This means that every active room (i.e., every instance that audio-content is generated on the app) has a “moderator” present (**mechanisms**). All other users that enter the room start out as listeners, and do not have the ability to speak in this role—they cannot unmute their microphone. As a listener, users can press the “raise hand” button and ask to be a speaker. If a moderator accepts a listener’s request to speak, that listener gets moved up to the “stage” where they now have the

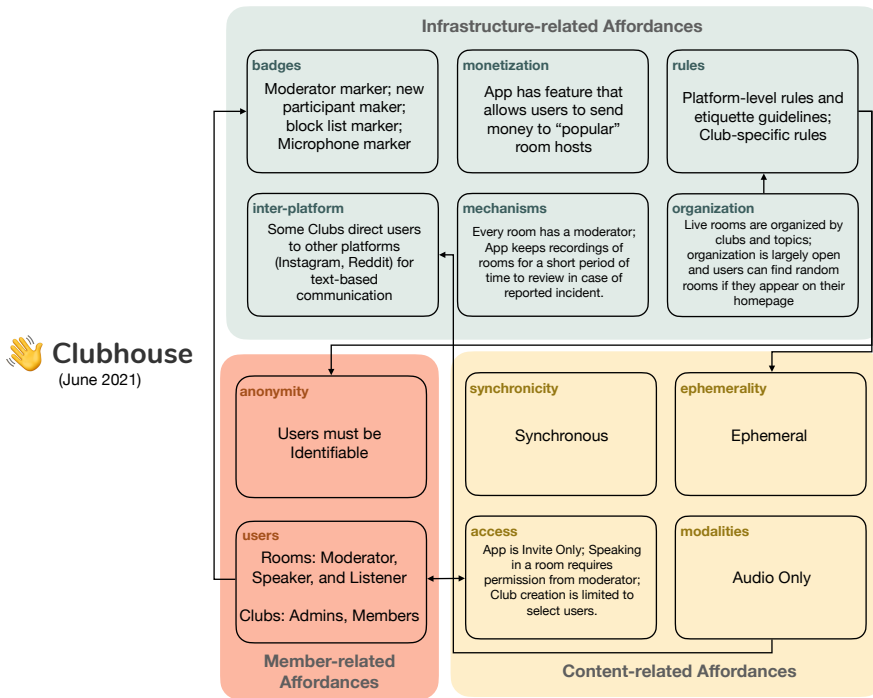


Fig. 4. MIC diagram for Clubhouse as of June 2021. At this point, Clubhouse users have no way of interacting with each other apart from using synchronous and ephemeral audio spaces (**modalities**); this led to many Clubhouse users using other platforms to engage in asynchronous text-based communication (**inter-platform**).

role of speaker. As a speaker, they can unmute their own microphone and be heard by everyone else in the room (**access**).

All speakers inside a room have a marker to show whether their microphone is muted or not. Speakers often click this marker on and off to indicate that they wish to speak next. When users enter a room, they have a celebratory emoji by their icon and name to indicate that they are new to the room (**badges**). Clubhouse also has a monetization feature that lets users send money to other Clubhouse users through their profile page (**monetization**). A block-list icon on profile pages are used to indicate to a user that the account has been blocked by many people in their circle (**mechanisms, badges**).

Text-based commentary or discourse pertaining to the interactions that occur on Clubhouse often happens on other platforms. One such platform that is heavily used to discuss particular events or incidents that happen in Clubhouse rooms is Twitter. Users will often talk about what they are experiencing on Clubhouse on Twitter, and Clubhouse users will often link to their Twitter profiles on their Clubhouse profile. There are also subreddits dedicated to talking about Clubhouse (i.e., r/Clubhouse). These other platforms are also used to announce and publicize rooms or clubs and invite new users to Clubhouse (**inter-platform**).

Moderation-Related Updates to Clubhouse. Between June of 2021 and January of 2022, Clubhouse released close to 20 updates to their iOS app [1]. These releases included changes to the app's appearance, updates to the app's terms of service and privacy policy, as well as the addition of

moderators to report disruptive or antisocial users. Clubhouse's record feature might allow room hosts to maintain records of users that engaged in disruptive behavior, as well as evidence of such behavior (**ephemerality, synchronicity**).

Before Clubhouse added a text-based chat feature, users had to utilize other social platforms if they wanted to send asynchronous, text-based messages to other users. This led to instances where abusive users used several other platforms to harass individuals they initially encountered on Clubhouse [30]. This type of behavior could amplify the amount of harassment a potential victim receives. Thus, the introduction of text-based messages (**modalities**) could likely reduce the reliance on these inter-platform relationships, making Clubhouse, and anti-social behavior that occurs on Clubhouse, more self-contained. This could potentially prevent the amplification of harassment that victims of antisocial users get. Finally, since Clubhouse is no longer invite-only (**access**), the user base of Clubhouse has undoubtedly expanded. This means more users, and more communities, will start using Clubhouse, likely resulting in a large influx of user and incident reports, thereby posing newer challenges to the platform.

5 MIC AS A TOOL FOR CROSS-PLATFORM MODERATION ANALYSIS

So far, we have created MIC diagrams for three platforms, all of which are centered around audio. As discussed in the introduction, these audio platforms have many similarities and differences that could impact how moderation is accomplished. In this section, we will compare and contrast the platforms via the MIC framework. Then, we will use these comparisons to generate ideas for new moderation interventions.

5.1 Similarities and Differences between Discord, Spotify, and Clubhouse

We begin by pointing out the obvious similarities and differences between the three platforms that can be determined without using MIC. First, Discord and Clubhouse both offer live audio features, whereas Spotify itself does not. Spotify also does not offer users a way to direct-message other users, while Discord and Clubhouse both have such features. In fact, Spotify users have no means to interact with one another on the platform apart from using posted audio, which is not the case on Discord or Clubhouse. In general, Spotify is used for listening to Music and Podcasts; Clubhouse is used for listening to and participating in live audio rooms; Discord is used to host communities and let community members interact with each other over text, voice, and video.

Affordance-based comparisons between Discord, Spotify, and Clubhouse are shown Table 1. Similarities and differences between inter-affordance relationships of the platforms can be seen by comparing the edges of the MIC diagrams found in Figures 2, 3 and 5.

5.2 Adapting and Proposing Moderation Mechanisms using MIC Comparisons

Spotify and Clubhouse. One challenge we noticed while using Clubhouse to conduct the previous case study (Section 4) is that it was difficult to identify live rooms that are of interest that appear on the app's home page. Furthermore, some live rooms dealt with sensitive topics, such as sexual assault. Such rooms should likely not be recommended or shown to users who are insensitive to certain topics, since their participation in the room would have negative impacts on the members of such a space. In general, it seems difficult for both listeners to find relevant and interesting rooms on Clubhouse and room hosts to find interested listeners and participants. To begin addressing this potential challenge, one can use MIC-based comparisons to observe that Clubhouse has a similar open **organization** to Spotify. In particular, the room topic categories that users can browse on Clubhouse are reminiscent of the various categories users can use to browse content on Spotify. Likewise, both platforms host non-ephemeral content (**ephemerality**).

| | Discord | Spotify | Clubhouse |
|-----------------------|---|---|--|
| modalities | Text, audio, and video | Primarily audio | Primarily audio |
| access | All users can generate content | Not all users can generate content | All users can generate content |
| monetization | No monetization tools for users | Content is monetized (through ads and stream count) | Users can be monetized (other users can donate to user profiles) |
| synchronicity | Synchronous (audio/video) and Asynchronous (text) | Asynchronous (audio) | Synchronous (audio) |
| ephemerality | Ephemeral (audio/video) and Non-ephemeral (text) | Non-ephemeral (audio) | Ephemeral (audio) and Non-ephemeral (audio/text) |
| organization | All communities organized into/within servers; Users can search for public servers using keyword search | Open organization; content organized by genre, artist, playlists etc. | Some communities organized into/within clubs; Public clubs and rooms categorized by topic |
| users | Custom user roles can be created for each server | Fixed, platform-wide user roles (associated with a user's account) | Fixed user roles that can change within Clubs and rooms |
| anonymity | Pseudonymity allowed | Pseudonymity allowed | Users must be Identifiable |
| rules | Platform-wide and server-specific rules | Platform-wide rules | Platform-wide and club-specific rules |
| badges | Custom roles can have custom visual markers | Verified Artist accounts have "Blue Check" | Users in rooms have badges associated with their roles |
| inter-platform | Can host communities independently; Often used in tandem with other platforms | Cannot host communities independently; Integrated into other platforms | Can host communities independently; Often used in tandem with other platforms |
| mechanisms | Automated tools for user-driven moderation | Automated tools for platform-driven moderation (through content flagging, curation, and recommendation) | Platform-driven moderation involves recording all audio rooms; No automated tools for user-driven moderation |

Table 1. MIC-guided comparisons between the affordances of moderation ecosystems on Discord, Spotify, and Clubhouse (as of January 2022). These comparisons will guide us in generating ideas for new moderation interventions for each platform.

One of Spotify's major services is its recommendation system for music and podcast discovery. Not only does this service aim to show users content that they would be inclined to listen to, but also for creators to discover new listeners.⁴ One way in which Spotify does this is by curating playlists. These playlists can be broadly defined, containing music from a genre, or from a specific musical artist. Many of these playlists are manually curated, and artists can submit music for consideration to be added to these curated playlists.

Given Clubhouse and Spotify's organizational similarity, and the existence of non-ephemeral content, we could propose a moderation mechanisms for Clubhouse that involves adopting a similar type of recommendation-via-curation mechanism like Spotify, and manually curate endorsed playlists of recordings of quality room recordings. We could even try to extend this idea to ephemeral content, i.e. playlist-type hubs of clubs or upcoming scheduled rooms that are hosted by trusted or experience users. This could start to help clubs and rooms find relevant audiences, and could also help users find and build communities in a more strategic way, while limiting the number of potential bad actors that try to engage.

Discord and Clubhouse. MIC also showed us that Clubhouse and Discord are very similar across many different affordances. Discord has been studied in the context of moderation research [25, 27], and researchers have found that moderating voice channels on Discord is a challenging feat. This is largely due to the fact that moderators in Discord servers find it difficult to monitor events and collect evidence of bad behavior in voice channels [25]. Clubhouse, like Discord, has a moderator role for **users**; however, on Clubhouse, every active room must have a moderator present. A feature, or moderation **mechanism**, that Discord could "borrow" from Clubhouse to help moderators handle voice-channels is a way to enable moderators to schedule when voice-channels can be made active. This way, moderators can ensure that they are present in public voice channels. Discord moderators can already limit when voice channels are open, but scheduling such times (similar to how live rooms are scheduled in Clubhouse clubs by Admins) can make this easier to do.

Another change Discord could make is to adopt Clubhouse's policy of keeping recordings of voice-rooms for a short period of time in order to address or investigate any reports (**rules**). It might be the case that some Discord servers have such a policy for their server; creating a platform-wide policy would be a more robust measure to discourage harmful behavior in such spaces. However, the pseudonymous nature of Discord (**anonymity**) might make such a policy not only difficult to implement, but also off-putting to Discord's user base (a large portion of which is comprised of gaming communities, where users have been shown to prefer some degree of anonymity and privacy [51]). Clubhouse's recording policy, when introduced, did not appear to drive away its user base. Though it is unclear exactly why this was the case, it could be because every user on the app has already agreed (in order to satisfy Clubhouse's Terms of Service) to be identifiable. Thus users have already agreed to forfeit some of their privacy. Clubhouse can adapt some moderation **mechanisms** from Discord as well. In particular, Clubhouse could develop an API or a collection of chat bots or tools that help to moderate text conversations. Such tools could also be developed for room moderators to help them keep track of members of a room, flag certain users, handle requests to speak, or manage music streams, as is the case with certain Discord bots [25]. It might be the case that different types of rooms or clubs want or need different types of tools, thus the customizability of Discord's moderation tools and API could be useful for Clubhouse users.

6 DISCUSSION

For CSCW theory, our framework provides a new analytic lens to identify, understand, and compare the various components of a social platform's moderation ecosystem. MIC allows moderation

⁴Both Anchor.fm and Spotify For Artists have tools for musicians to help them expand their audience.

researchers and stakeholders to efficiently and comprehensively navigate moderation-specific aspects of social platforms. The various insights MIC led us to can be used to develop research questions that moderation researchers can use to further investigate new and dynamic platforms like Clubhouse and motivate future studies. Likewise, platform designers and moderators themselves can use these insights to preemptively infer potential moderation challenges that might arise, and can prepare for them by designing new tools, features, or guidelines. Comparing moderation ecosystems across platforms using MIC can allow stakeholders to adopt successful moderation mechanisms from one another without overlooking subtle but potentially significant differences. We now discuss further implications, potential limitations, and extensions of MIC.

6.1 Implications and Advantages of Using MIC

Efficient Navigation of New Platforms. Platforms often offer a plethora of features which can make it difficult to discern which features are relevant for moderation. MIC allows us to systematically pinpoint the facets of a platform's design and affordances that are relevant. In our case studies, we used MIC to determine relevant features to examine their role in effecting moderation on different platforms. For instance, Clubhouse has other features that are not described in the previous two sections, since they do not fall under any of MIC's affordances. One such feature is Clubhouse's calendar page, which displays upcoming rooms that are scheduled for each user. We were not able to uncover ways in which the calendar feature enables anti-social behavior or promotes pro-social behavior, or aids in moderating the platform. As such, it is omitted in our representation using MIC, allowing us to focus on just the features that are relevant.

Accounting for Nuances at the Platform Level. Though Jiang et al. [25] uncovered the challenges that come with "real-time, ephemeral" voice communities, our case studies have shown that other platforms that also have **synchronous** and **ephemeral** audio might not face the same problems as those found in Discord. Namely, Jiang et al. [25] found that moderators on Discord have a tough time gathering evidence of antisocial behavior since they are not present in voice channels at all times; this creates challenges since it limits the moderators' ability to exclude bad actors, and exclusion is one of Grimmelmann's four techniques for moderation. Guided by MIC, we found that Clubhouse rooms always have moderators present (**users, mechanisms**). Thus, potential moderation challenges on Clubhouse may not be due to the lack of mechanisms for exclusion, as it appears to be on Discord. Alternatively, there could be other unidentified novel consequences of synchronous, ephemeral voice settings that hinders moderators' ability to exclude bad actors.

We believe that MIC-derived insights, which can be explored in future studies regarding moderation on Clubhouse or synchronous, ephemeral audio platforms at large, are far more comprehensive than those that could be derived from Grimmelmann's view of moderation [20]. In fact, Jiang et al. [25] ultimately argue that "designers and moderators should not ignore the technological infrastructure of the community" when trying to implement existing moderation strategies to new communities, and that they should "carefully consider the limitations" imposed by new infrastructure. Analysis via MIC begins with understanding the technological infrastructure of communities as it spells out the moderation-specific aspects of the platforms they are situated on. Having these insights could aid in future study design, and make the road to uncovering and extending findings, such as the ones from Jiang et al. [25], clearer.

Understanding how Platform Changes Affect Moderation. Another benefit of using MIC is that it helps us pinpoint how specific changes on a platform could impact moderation (Section 4). Furthermore, we were able to use the inter-affordance relationships identified in MIC to get a more complete understanding of potential ways in which certain updates could effect Clubhouse's moderation ecosystem. For example, Clubhouse's new text-based messaging feature caused us to

update the **modalities** affordance. However, since we used MIC to analyze Clubhouse, we observed that users used other platforms in tandem with Clubhouse to message one another (and therefore, a relationship between the **modalities** affordance and the **inter-platform** affordance). Thus, we could consider the possibility that a change to the **modalities** affordance would result in a change to **inter-platform** affordance. Using this relationship, we discussed potential impacts the above change might have had to moderation on Clubhouse. Without MIC, we might have overlooked this relationship and failed to investigate inter-platform relationships after modality changes.

Additionally, changes on Clubhouse occurred over a period of six months, which is as long as a revision cycle in publication venues like CSCW. This means that moderation research and proposed moderation tools may become out-dated or obsolete more quickly. Using MIC as a common foundation with which to discuss moderation on social platforms would allow us to easily adapt and discuss how changes and updates to a platform may impact results of research and design.

Adapting Moderation Mechanisms from Other Platforms. Section 5 demonstrates how MIC can be used to compare platforms in a systematic manner. MIC allows us to be mindful of how similar features across platforms can actually be impacted by different affordances. For instance, while Clubhouse and Spotify both have non-ephemeral content, Spotify's content is created asynchronously, while Clubhouse's content is created synchronously. Hence, while both platforms can moderate such content after-the-fact, Clubhouse has additional measures to ensure safety in live rooms. It is unclear to what additional moderation mechanisms Clubhouse has, if any, for its non-ephemeral content, apart from those listed in its Terms of Service. However, in comparing Spotify and Clubhouse, we could propose potential mechanisms for Clubhouse that are inspired by affordances in Spotify. Similarly, we used the comparisons between Discord and Clubhouse to propose moderation mechanisms for each platform that are inspired by each other.

Broader Effects of Affordance Changes. In Section 5.2, we discussed potential moderation mechanisms that Discord, Spotify, and Clubhouse could adapt from one another. One proposal we made involved adapting Clubhouse's rule of keeping all recordings for a short period of time to address voice moderation challenges found on Discord [25]. We briefly discussed that users of Discord may not be open to this platform change, largely due to the fact that Discord seems to allow its users more privacy than Clubhouse does. This conjecture was made by observing that Discord users are allowed to be pseudonymous, while Clubhouse users have always been required to be identifiable. Observations like this are seemingly unimportant, and had we not used MIC, may have been overlooked. However, in some cases, design decisions that overlooked these subtle nuances directly preceded the downfall of platforms.

An example of this can be seen with YikYak, a social platform that allowed users to post location-specific anonymous text-posts [41]. YikYak was a successful social platform that shut down in 2017 after platform changes were introduced. One such update was the removal of anonymity. As discussed in Section 2, existing research has explored the role anonymity played in voice-based interactions in online games [51]. In particular, Wadley et al. [51] found that voice seemed to remove a degree of anonymity in game-play, which made some players feel uncomfortable, and in some cases, caused the players to abandon the game. We cannot retroactively prove that MIC-based analysis would have prevented platform designers from making this platform change, nor can we assert that this specific change was the reason users abandoned the platform. However, MIC would have highlighted anonymity as an integral affordance, and one that was similar to that of the online games and gaming platforms explored by Wadley et al. [51]. MIC-based analysis could have made these connections to a seemingly unrelated platform clear, and could have shed light on potential challenges that could result from altering the platform's anonymity affordance. As such, MIC-based

approach to moderation research and social platform design could be a valuable tool in designing and maintaining successful social platforms.

6.2 Limitations and Future Work

Limitations of MIC. MIC's purpose is for capturing the moderation ecosystems of social platforms to allow moderation researchers and platform stakeholders to better understand moderation. However, MIC does not capture every moderation-related property. In particular, the implicit norms that exist on a platform would not be represented by the affordances or relationships in MIC, since they are not tangible. Norms of online communities play a massive role in moderation on platforms, and is identified as one of four main moderation techniques by Grimmelmann [20]; there is also research that explores how norms play a role in moderating online communities, and how norms differ amongst various communities on the same platform [14, 42].

Another closely related limitation of MIC is that it is not currently designed for analyzing the individual communities on each social platform. Studying individual online communities, such as specific subreddits, is beneficial for understanding moderation since each community has its own unique norms [17]. There might be a way to extend MIC to capture nuances of community norms, which could be explored in future work.

Extending MIC. MIC's base set of affordances and relationships are likely to become non-exhaustive as technology advances. Luckily, the graphical nature of MIC allows us to do so in an easy and straightforward way. We can add new affordances to our original set when new types of affordances that effect moderation are uncovered or developed. Similarly, we could further granularize existing affordances. For instance, we may eventually find it useful to distinguish between automated moderation mechanisms and manual ones. We can also extend our set of relationships by defining new types of relationships. There is no real restriction on how one could go about defining new relationships. We could even forego the condition that relationships occur between only two affordances, and describe multi-affordance relationships that are analogous to *hyper-edges*.⁵

Another potentially useful, albeit more involved, extension of MIC, and in particular the MIC diagram would be to use the inter-platform relationship affordance with a MIC diagram for other platforms or services. This would be useful if there is a nearly symbiotic relationship between two separate platforms or services, but we still wish to consider the affordances of each separately. For instance, Discord developed a new Clubhouse-like addition called Discord Stages⁶. It may be useful to consider Stages as a separate service from Discord's servers, since its use-case and set-up is different. We could analyze each of these services separately, and then build an extended MIC diagram to understand moderation on Discord in more detail.

7 CONCLUSION

In this paper, we introduced the MIC framework as an affordance-aware augmentation of Grimmelmann's taxonomy [20], a popular lens for discussing moderation. MIC provides a standardized way to represent moderation ecosystems of social platforms that highlights the moderation-related platform affordances involved, as well as the relationships that exist between them. Over the course of two case studies, we used MIC to analyze a rapidly evolving platform (Clubhouse) and subsequently compare it to other relevant platforms (Discord and Spotify) to help generate possible new moderation interventions to address the challenges each may face. We believe that the dynamic

⁵In graph theory, a hyper-edge represents an "edge" that occurs between more than two vertices, usually represented as a subset of vertices.

⁶<https://discord.com/stages>

and comprehensive nature of the MIC framework will help the moderation research community and stakeholders effectively keep up with the fast-paced nature of social platform development.

REFERENCES

- [1] [n.d.]. Clubhouse iOS Release Notes. <https://www.notion.so/iOS-Release-Notes-acfb2f5d56cf4718b6486f5f670db6ad>
- [2] 2021. Community Guidelines. <https://www.notion.so/Community-Guidelines-461a6860abda41649e17c34dc1dd4b5f>
- [3] 2021. Discord introduces Clubhouse-like Stage Channels feature for live audio conversations- Technology News, Firstpost. <https://www.firstpost.com/tech/news-analysis/discord-introduces-clubhouse-like-stage-channels-feature-for-live-audio-conversations-9490401.html>
- [4] 2021. Spotify Acquires Locker Room and Announces Plans for a New Live Audio Experience. <https://newsroom.spotify.com/2021-03-30/spotify-acquires-locker-room-and-announces-plans-for-a-new-live-audio-experience/>
- [5] Mark S. Ackerman, Brian Starr, Debby Hindus, and Scott D. Mainwaring. 1997. Hanging on the ‘Wire: A Field Study of an Audio-Only Media Space. *ACM Trans. Comput.-Hum. Interact.* 4, 1 (March 1997), 39–66. <https://doi.org/10.1145/244754.244756>
- [6] Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. 2013. Steering User Behavior with Badges. In *Proceedings of the 22nd International Conference on World Wide Web (Rio de Janeiro, Brazil) (WWW ’13)*. Association for Computing Machinery, New York, NY, USA, 95–106. <https://doi.org/10.1145/2488388.2488398>
- [7] Michael Bernstein, Andrés Monroy-Hernández, Drew Harry, Paul André, Katrina Panovich, and Greg Vargas. 2011. 4chan and/b: An Analysis of Anonymity and Ephemerality in a Large Online Community. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 5.
- [8] Thales Bertaglia, Adrien Dubois, and Catalina Goanta. 2021. Clout Chasing for the Sake of Content Monetization: Gaming Algorithmic Architectures with Self-moderation Strategies. *Forthcoming in Morals and Machines 1* (2021).
- [9] Lindsay Blackwell, Nicole Ellison, Natasha Elliott-Deflo, and Raz Schwartz. 2019. Harassment in Social Virtual Reality: Challenges for Platform Governance. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 100 (Nov. 2019), 25 pages. <https://doi.org/10.1145/3359202>
- [10] Ragnhild Brøvig-Hanssen and Ellis Jones. 0. Remix’s retreat? Content moderation, copyright law and mashup music. *New Media & Society* 0, 0 (0), 14614448211026059. <https://doi.org/10.1177/14614448211026059> arXiv:<https://doi.org/10.1177/14614448211026059>
- [11] Ashley Carman. 2021. Spotify is launching its own Clubhouse competitor. <https://www.theverge.com/2021/3/30/22356993/spotify-locker-room-clubhouse-launch-acquisition>
- [12] Eshwar Chandrasekharan, Chaitrali Gandhi, Matthew Wortley Mustelier, and Eric Gilbert. 2019. Crossmod: A Cross-Community Learning-Based System to Assist Reddit Moderators. 3, CSCW (2019). <https://doi.org/10.1145/3359276>
- [13] Eshwar Chandrasekharan, Umashanthi Pavalanathan, Anirudh Srinivasan, Adam Glynn, Jacob Eisenstein, and Eric Gilbert. 2017. You Can’t Stay Here: The Efficacy of Reddit’s 2015 Ban Examined Through Hate Speech. 1, CSCW (2017). <https://doi.org/10.1145/3134666>
- [14] Eshwar Chandrasekharan, Mattia Samory, Shagun Jhaver, Hunter Charvat, Amy Bruckman, Cliff Lampe, Jacob Eisenstein, and Eric Gilbert. 2018. The Internet’s Hidden Rules: An Empirical Study of Reddit Norm Violations at Micro, Meso, and Macro Scales. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 32 (nov 2018), 25 pages. <https://doi.org/10.1145/3274301>
- [15] Kimberly M. Christopherson. 2007. The positive and negative implications of anonymity in Internet social interactions: “On the Internet, Nobody Knows You’re a Dog”. *Computers in Human Behavior* 23, 6 (2007), 3038–3056. <https://doi.org/10.1016/j.chb.2006.09.001> Including the Special Issue: Education and Pedagogy with Learning Objects and Learning Designs.
- [16] Robert B. Cialdini and Melanie R. Trost. 1998. *Social influence: Social norms, conformity and compliance*. McGraw-Hill, New York, NY, US, 151–192.
- [17] Sarah Gilbert. 2020. “I run the world’s largest historical outreach project and it’s on a cesspool of a website.” Moderating a Public Scholarship Site on Reddit: A Case Study of r/AskHistorians. *Proceedings of the ACM on Human-Computer Interaction* 4 (05 2020), 1–27. <https://doi.org/10.1145/3392822>
- [18] Tarleton Gillespie. 2018. *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.
- [19] Robert Gorwa, Reuben Binns, and Christian Katzenbach. 2020. Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society* 7, 1 (2020), 2053951719897945. <https://doi.org/10.1177/2053951719897945> arXiv:<https://doi.org/10.1177/2053951719897945>
- [20] James Grimmelmann. 2015. The virtues of moderation. *Yale JL & Tech.* 17 (2015), 42.
- [21] Stephen C. Hayne and Ronald E. Rice. 1997. Attribution accuracy when using anonymity in group support systems. *International Journal of Human-Computer Studies* 47, 3 (1997), 429–452. <https://doi.org/10.1006/ijhc.1997.0134>

- [22] Ellen A. Isaacs and John C. Tang. 1993. What Video Can and Can't Do for Collaboration: A Case Study. In *Proceedings of the First ACM International Conference on Multimedia* (Anaheim, California, USA) (MULTIMEDIA '93). Association for Computing Machinery, New York, NY, USA, 199–206. <https://doi.org/10.1145/166266.166289>
- [23] Shagun Jhaver, Iris Birman, Eric Gilbert, and Amy Bruckman. 2019. Human-Machine Collaboration for Content Regulation: The Case of Reddit Automoderator. *ACM Trans. Comput.-Hum. Interact.* 26, 5, Article 31 (July 2019), 35 pages. <https://doi.org/10.1145/3338243>
- [24] Shagun Jhaver, Sucheta Ghoshal, Amy Bruckman, and Eric Gilbert. 2018. Online Harassment and Content Moderation: The Case of Blocklists. *ACM Trans. Comput.-Hum. Interact.* 25, 2, Article 12 (March 2018), 33 pages. <https://doi.org/10.1145/3185593>
- [25] Jialun Aaron Jiang, Charles Kiene, Skyler Middler, Jed R. Brubaker, and Casey Fiesler. 2019. Moderation Challenges in Voice-Based Online Communities on Discord. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 55 (Nov. 2019), 23 pages. <https://doi.org/10.1145/3359157>
- [26] Charles Kiene, Jialun Aaron Jiang, and Benjamin Mako Hill. 2019. Technological Frames and User Innovation: Exploring Technological Change in Community Moderation Teams. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 44 (Nov. 2019), 23 pages. <https://doi.org/10.1145/3359146>
- [27] Charles Kiene, Andrés Monroy-Hernández, and Benjamin Mako Hill. 2016. Surviving an "Eternal September": How an Online Community Managed a Surge of Newcomers. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) (CHI '16). Association for Computing Machinery, New York, NY, USA, 1152–1156. <https://doi.org/10.1145/2858036.2858356>
- [28] Sara Kiesler, Robert Kraut, Paul Resnick, and Aniket Kittur. 2012. Regulating behavior in online communities. *Building successful online communities: Evidence-based social design* (2012), 125–178.
- [29] Cliff Lampe and Paul Resnick. 2004. *Slash(Dot) and Burn: Distributed Moderation in a Large Online Conversation Space*. Association for Computing Machinery, New York, NY, USA, 543–550. <https://doi.org/10.1145/985692.985761>
- [30] Taylor Lorenz. 2021. Clubhouse Moderation Issues and Incidents. <https://taylorlorenz.medium.com/clubhouse-moderation-issues-and-incidents-726a88a1b4bd>
- [31] J. Nathan Matias. 2016. Going Dark: Social Factors in Collective Action Against Platform Operators in the Reddit Blackout. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) (CHI '16). Association for Computing Machinery, New York, NY, USA, 1138–1151. <https://doi.org/10.1145/2858036.2858391>
- [32] J. Nathan Matias. 2019. The Civic Labor of Volunteer Moderators Online. *Social Media + Society* 5, 2 (2019), 2056305119836778. <https://doi.org/10.1177/2056305119836778> arXiv:<https://doi.org/10.1177/2056305119836778>
- [33] Poppy Laretta McLeod. 1997. A Comprehensive Model of Anonymity in Computer-Supported Group Decision Making. In *Proceedings of the Eighteenth International Conference on Information Systems* (Atlanta, Georgia, USA) (ICIS '97). Association for Information Systems, USA, 223–234.
- [34] Anne Oeldorf-Hirsch and German Neubaum. 2021. What Do We Know about Algorithmic Literacy? The Status Quo and a Research Agenda for a Growing Field. (2021).
- [35] Judith S Olson, Gary M Olson, and David K Meader. 1995. What mix of video and audio is useful for small groups doing remote real-time design work?. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 362–368.
- [36] Neil Patel, Deepti Chittamuru, Anupam Jain, Paresh Dave, and Tapan S. Parikh. 2010. *Avaaj Otalo: A Field Study of an Interactive Voice Forum for Small Farmers in Rural India*. Association for Computing Machinery, New York, NY, USA, 733–742. <https://doi.org/10.1145/1753326.1753434>
- [37] Jay Peters. 2021. Reddit Talk is a Clubhouse competitor for subreddits. <https://www.theverge.com/platform/amp/2021/4/19/22391875/reddit-talk-clubhouse-social-audio>
- [38] Damian Radcliffe. 2021. Audio Chatrooms like Clubhouse Have Become the Hot New Media by Tapping into the Age-Old Appeal of the Human Voice. *The Conversation* (2021).
- [39] Sarah T Roberts. 2016. Commercial content moderation: Digital laborers' dirty work. (2016).
- [40] Nadeem Sarwar. 2021. Slack is getting Clubhouse-like audio chatrooms, and I absolutely don't want it! <https://pocketnow.com/slack-is-getting-clubhouse-like-audio-chatrooms-and-i-absolutely-dont-want-it>
- [41] Ari Schlesinger, Eshwar Chandrasekharan, Christina A. Masden, Amy S. Bruckman, W. Keith Edwards, and Rebecca E. Grinter. 2017. *Situated Anonymity: Impacts of Anonymity, Ephemerality, and Hyper-Locality on Social Media*. Association for Computing Machinery, New York, NY, USA, 6912–6924. <https://doi.org/10.1145/3025453.3025682>
- [42] Joseph Seering, Robert Kraut, and Laura Dabbish. 2017. Shaping Pro and Anti-Social Behavior on Twitch Through Moderation and Example-Setting. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) (CSCW '17). Association for Computing Machinery, New York, NY, USA, 111–125. <https://doi.org/10.1145/2998181.2998277>

- [43] Joseph Seering, Tony Wang, Jina Yoon, and Geoff Kaufman. 2019. Moderator engagement and community development in the age of algorithms. *New Media & Society* 21, 7 (2019), 1417–1443. <https://doi.org/10.1177/1461444818821316> arXiv:<https://doi.org/10.1177/1461444818821316>
- [44] Julius Solans. 2020. The Rise of Audio in Virtual Events to Combat Zoom Burnout. (2020).
- [45] Inc. Sonar. 2021. Sonar: Create worlds together. <https://apps.apple.com/us/app/sonar-create-worlds-together/id1512829586>
- [46] Anne Steele. 2021. Spotify Acquires Sports-Talk App Locker Room. <https://www.wsj.com/articles/spotify-acquires-sports-talk-app-locker-room-11617109208>
- [47] Charlotte Tang and Sheelagh Carpendale. 2009. *A Mobile Voice Communication System in Medical Setting: Love It or Hate It?* Association for Computing Machinery, New York, NY, USA, 2041–2050. <https://doi.org/10.1145/1518701.1519012>
- [48] Emily Tarvin and Mel Stanfill. 2022. “YouTube’s predator problem”: Platform moderation as governance-washing, and user resistance. *Convergence* (2022), 13548565211066490.
- [49] Harry C. Triandis. 2004. *Culture and social behavior*. McGraw-Hill, Custom Publishing.
- [50] Aditya Vashistha, Edward Cutrell, Gaetano Borriello, and William Thies. 2015. *Sangeet Swara: A Community-Moderated Voice Forum in Rural India*. Association for Computing Machinery, New York, NY, USA, 417–426. <https://doi.org/10.1145/2702123.2702191>
- [51] Greg Wadley, Marcus Carter, and Martin Gibbs. 2015. Voice in Virtual Worlds: The Design, Use, and Influence of Voice Chat in Online Play. *Human–Computer Interaction* 30, 3-4 (2015), 336–365. <https://doi.org/10.1080/07370024.2014.987346>
- [52] Suzanne Weisband. 1993. Overcoming social awareness in computer-supported groups. *Computer Supported Cooperative Work (CSCW)* 2, 4 (01 Dec 1993), 285–297. <https://doi.org/10.1007/BF00805695>
- [53] Sarah Myers West. 2018. Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. *New Media & Society* 20, 11 (2018), 4366–4383. <https://doi.org/10.1177/1461444818773059>
- [54] Luke Wilson. 2021. Telegram is stealing the best feature from Clubhouse - here’s how. <https://www.tomsguide.com/news/telegram-is-stealing-the-best-feature-from-clubhouse-heres-how>
- [55] Donghee Yvette Wohn. 2019. *Volunteer Moderators in Twitch Micro Communities: How They Get Involved, the Roles They Play, and the Emotional Labor They Experience*. Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3290605.3300390>
- [56] Bin Xu, Pamara Chang, Christopher L. Welker, Natalya N. Bazarova, and Dan Cosley. 2016. Automatic Archiving versus Default Deletion: What Snapchat Tells Us About Ephemerality in Design. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (San Francisco, California, USA) (CSCW ’16). Association for Computing Machinery, New York, NY, USA, 1662–1675. <https://doi.org/10.1145/2818048.2819948>

Received July 2021; revised November 2021; accepted April 2022