Values (Mis)alignment: Exploring Tensions Between Platform and LGBTQ+ Community Design Values

MICHAEL ANN DEVITO, University of Colorado Boulder, USA
ASHLEY MARIE WALKER, Northwestern University, USA
JULIA R. FERNANDEZ, Northwestern University, USA

Social platforms hold great promise for supporting marginalized communities, such as the LGBTQ+ community, yet they are frequently sites of further stigmatization and harm. By engaging a diverse sample of 31 US LGBTQ+ users in five qualitative, design-based value elicitation exercises, we find that misalignments between perceived platform values and the values of the marginalized users they serve are at the heart of this disconnect. We inductively identify two community-based design values for supporting LGBTQ+ users: self-determination and inclusion. These values can be used as design heuristics for both improving existing platforms as well as guiding future new platform development. Based on participant feedback, we provide directions for enacting these values to better align platform values with this marginalized population's needs.

CCS Concepts: • **Social and professional topics~Gender** • **Social and professional topics~Sexual orientation** • Human-centered computing~Social networking sites • Human-centered computing~Empirical studies in HCI

KEYWORDS: Design values, value-sensitive design, online communities, LGBTQ+, Queer HCI, social media, platforms, anxiety

ACM Reference format:

Michael Ann DeVito, Ashley Marie Walker, and Julia R. Fernandez. 2021. Values (Mis)alignment: Exploring Tensions Between Platform and LGBTQ+ Community Design Values. In *Proceedings of the ACM on Human-Computer Interaction,* Vol. 5, CSCW1, Article 88 (April 2021), 27 pages, https://doi.org/10.1145/3449162

1 INTRODUCTION

When well-constructed and well-managed, online social platforms (e.g., Facebook, Twitter) have the potential to be positive forces for users from marginalized groups. Social platforms can be instruments of not just social, informational, health, and identity-development support (e.g., as they are for the LGBTQ+ community [5, 16, 29, 45, 48, 86]), but also community solidarity and personal empowerment [65]. However, the current state of social platforms does not always fulfill this promise for marginalized users and may, in fact, harm them. For the LGBTQ+ community, social platforms continue to be a source of bullying and harassment [37], and have been central

This work is supported by the Sexualities Project at Northwestern University.

Authors' addresses: Michael Ann DeVito, University of Colorado Boulder, Department of Information Science, 1045 18thSt., Boulder, CO, 80309, USA, michaelann@colorado.edu; Ashley Marie Walker & Julia R. Fernandez, Northwestern University School of Communication, 2240 Campus Drive, Evanston, IL, USA, 60208, amwalker@u.northwestern.edu & j-fernandez@u.northwestern.edu.

https://doi.org/10.1145/3449162

to incidents that harm LGTBQ+ people and members of other marginalized groups (e.g., Gamergate [64]). Structurally, the overall platform emphasis on growth and engagement metrics also results in the replication of problematic power structures from the offline world in the online environment, including structures that allow and even support the intracommunity stigmatization and harassment of transgender and non-monosexual[1] individuals [74, 83]). Even when trying to solve problems that negatively impact LGBTQ+ people, such as the ongoing crisis in content moderation [36], the resulting platform solutions, in the form of new policies and algorithms, may unduly censor LGBTQ+ voices [4, 15]. Social platforms could realize substantial benefits for LGBTQ+ people. However, due to the fact that they are rarely designed *for* [1, 41] and even more rarely *with* LGBTQ+ people [68], platforms have an inadequate or incorrect understanding of the needs of this marginalized group.

One approach to understanding these needs is to engage the LGBTQ+ community through the lens of user values. HCI has a history of values as the basis for technology design [31, 32] that address the needs of marginalized communities (e.g., [10, 19, 54, 56, 88]). Values inform user decision-making about platform use/nonuse and participation [11, 75], as well as user understanding/trust in the algorithmic components of platforms [71], making them key practical concerns for developers. Values-based approaches also help reflect the diversity of the community under study, allowing for input from a wide range of non-technical community stakeholders [82]. This diversity provides an opportunity to incorporate a plurality of perspectives and wrestle with tensions that may exist within marginalized communities [30]. In the specific context of the LGBTQ+ community, this is crucial to surface values which balance the intracommunity solidarity necessary for dealing with powerful platforms with the variance in identities and concerns within the LGBTQ+ community [23, 34]. Similarly, values exist at a high enough level that they can apply broadly, beyond one platform at one moment in time. This is essential for representing the LGBTQ+ community, as most LGBTQ+ people rely on a multi-platform ecosystem [20], emphasizing the importance of both improving existing platforms and building novel platforms to different subgroups [47].

In order to aid platforms in realizing their potential as crucial sites of information, social support, and empowerment for marginalized groups, this paper asks:

What values do a diverse sample of members of the LGBTQ+ community find important to attend to in social platform design?

Accordingly, we convened an online community with a diverse sample of 31 US LGBTQ+ social technology users. We engaged them in a series of five qualitative, value elicitation activities inspired by prior work in value-sensitive, participatory, and user-centered design (e.g., [3, 6, 30, 32, 57, 82, 85]). Through an adaptive, five-stage study design (positive/negative experiences, blue sky needs/wants, practical algorithmic/automatic concerns, practical moderation/policy concerns, prioritization/role-taking), we found that LGBTQ+ users see utility in social platforms, but are anxious about the current and future state of these platforms. This anxiety is caused by a misalignment between LGBTQ+ community values and what LGBTQ+ users perceive as platform values based on their interactions with the platforms. In contrast to perceived platform values, we contribute two design-relevant values based in the lived experiences [56] of LGBTQ+ social platform users: *self-determination* and *inclusion*. Both values have implications for platform design, algorithm design, and moderation tools/policy. We present these values, and our participants' suggestions for enacting them through design while managing both logistical and intracommunity trade-offs, as a step towards better aligning current platforms with community needs for marginalized groups while providing heuristics for future design work.

[1] Individuals who are attracted to more than one gender, e.g. bisexuals and pansexuals.

2 BACKGROUND

Research on online communities ranges from Kraut and Resnick's work on how to build and maintain online communities [55], to work on how communities cope with sudden change [51], and work on free and open-source software (FOSS) collaborations, such as Wikipedia (e.g. [44, 78]). This work often has implicit values embedded within it. For example, Kraut, Resnick, and their collaborators prioritize community growth, user retention, and boosting user contributions [55], while FOSS work tends to prioritize growth alongside quality of contributions in relation to the goal of making a good product [44, 78]. Social platforms also have their own goals, including platform growth, boosting content engagement, and avoiding legal trouble [35, 36]. In all cases, these goals exist across a broad base of potential users - anyone, for example, can become a Wikipedia editor so long as they follow the rules, and participate in creating an accurate, free encyclopedia.

Marginalized groups, however, may not be responsive to these goals, and organize around different principles. Online communities for LGBTQ+ people are sites of identity development, informational, and social support [5, 16, 25, 29, 37, 38, 59, 86], and may have goals that are more about solidarity and empowerment than growth or product creation [65]. Additionally, these groups must deal with the challenges of organizing around (and disclosing) a potentially-stigmatized identity [2, 12, 20], and the attendant challenges of dealing with tensions among subgroups within that identity (e.g. [23, 34, 74, 83]).

2.1 Value Elicitation and Marginalized Populations

To better understand the needs and values of LGBTQ+ users, we draw from the value-sensitive design (VSD) tradition. Value-sensitive design encourages us to explore user values, which in this context refers to the contours of what users find important in the design of systems, integrating the user's actual experience of a system with views of how a system "ought to" work [31, 32]. In its original formulation, VSD is a design method, a three-part framework which steps through conceptual (literature and philosophy-centric), empirical (user-centric), and technical (product-centric) investigations of values and how they relate to users values for product development [31, 32].

Subsequent critiques, such as those by Borning and Muller [6] and Le Dantec et al. [56] have advocated for refocusing VSD away from a canonical set of values and instead prioritizing the local values of the community under study. Borning and Muller point to the utility of participatory design-influenced methods for this purpose [6], as participatory design traditions encourage engaging with stakeholders, who are experts in the sociotechnical context in which the system operates, to clarify goals and needs for information systems [77]. In the same vein, Le Dantec et al. advocate arranging VSD to prioritize the empirical investigation over conceptual or technical investigations of values, ensuring the values are community-based [56].

This community-first, empiricism-first variant of VSD has been used to elicit values from marginalized groups. For example, Koepfler et al. used empirically-led VSD to identify values for supporting homeless users from Twitter data and survey responses [54], while Zolyomi et al. used this process to identify design values for supporting autistic members of neurodiverse teams in education [88]. This process has also been used to generate design recommendations for situations where current systems do not appear to serve the purposes of the group under study. Deng et al. used a VSD approach to highlight the values of workers on Mechanical Turk and their disconnect from mTurk platform's values [19]. Similarly, Briggs and Thomas identified and provided solutions for value disconnects between marginalized people and the technologies through which they express their marginalized identities [10]. In these cases, directly engaging stakeholders in value elicitation based in their own lived experiences revealed not only the shape of the problem facing the marginalized participants, but also values which could help designers address these problems.

Accordingly, we adopt VSD as our overall framework here. Per Le Dantec et al., we focus on empirical investigation to prioritize the values of the LGBTQ+ community over the current values of the platform or a universal set of values [56]. Our hope is that this will enable future versions of platforms to be responsive to these values. We take our initial cues from Friedman et al., who suggest starting with a value, technology, or "context of use" [32], and focus on the intersection of a technology (social platforms) and a context of use (the functions of LGBTQ+ online communities).

2.2 A Primer on the LGBTQ+ Experience of Social Platforms

LGBTQ+ people, like most marginalized groups, derive essential benefits from the internet under the right circumstances [65]. In light of the decline in offline LGBTQ+ spaces [22], the internet plays an outsized role in LGBTQ+ socialization [40], and social platforms have become the seat of multiple functions for these individuals. This includes the exploration, development, and claiming of one's identity as an LGBTQ+ person [16, 29, 45]. It also includes access to social networks of support, connection, and information for LGBTQ+ people and their families [5, 25, 38, 49], including health information [48, 59]. Access to LGBTQ+ spaces on social platforms results in positive health outcomes, increased resistance to victimization, and friendships with other LGBTQ+ people among LGBTQ+ youth [37, 86].

However, platforms are not often designed by prioritizing the needs of the marginalized. For example, design principles may reflect biases such as publicness as a default, creating situations where platforms cause context collapse to the point of inadvertently outing users [14]. Similar risks exist where it is difficult to know if one is in a safe or unsafe space for seeking information or support. [25]. Ultimately, we rarely design systems with LGBTQ+ people in mind [1, 41], and even less frequently involve LGBTQ+ people as designers [68], leading to systems which further harm and marginalize people [1]. Further, this dynamic also fails to achieve the platform's own goals of utility for their user base [53]. While social platforms are a valuable tool for LGBTQ+ users, there is much room for improvement in orienting the values embedded within design to these needs, and this improvement requires the direct involvement of LGBTQ+ people who are experts on their own values and needs [68].

2.3 Diversity and Solidarity Within the LGBTQ+ Community: A Design Challenge

In designing to support LGBTQ+ people, we must also account for the fact that the LGBTQ+ community has substantial internal diversity [8]. Practically, this has resulted in systems designed for only certain subgroups of the LGBTQ+ community. For example, the design of geospatial networking platforms such as Grindr are not a one-size-fits-all solution for the entire LGBTQ+ community, and may perpetuate feelings of isolation among some users [24, 46]. The framing of LGBTQ+ spaces on social platforms as "safe spaces" is contradicted by evidence that social structures and policies do not prevent the infiltration of bad actors or the presence of intracommunity harm towards subgroups such as the transgender community or non-monosexual people [74, 83]. Indeed, in their identification of online threat models, Lerner et al. note that other LGBTQ+ people are explicitly part of the threat models of transgender individuals [58].

This internal diversity has led to debate over if LGBTQ+ people are, in fact, one community, representing a collective identity – e.g. one large "queer" community – instead of loosely related gay, lesbian, bisexual, etc., communities [34]. In a platform context, this might prompt one to ask "why not have separate spaces/rules for each subgroup?" Indeed, studies of individual subgroups such as transgender/nonbinary [42, 74] and bi+ individuals [83] have been crucial in identifying problems and design solutions specific to these doubly-marginalized groups [84]. This mirrors the larger community debate, where an initial focus on a tightly-defined, collective gay/lesbian

identity [34, 87] has been countered with attempts at queering, or troubling, that collective identity to better recognize the internal diversity of the community, and especially the presence of groups which still experience intracommunity stigmatization (e.g., transgender and bisexual people) [23, 34]. However, as queering in this context inherently involves challenging and potentially modifying or abandoning previously-stable concepts, it poses a threat to the stability of a collective identity [23, 34].

As LGBTQ+ people remain a marginalized group, there are good reasons to maintain a strong collective identity. In general, collective identities are crucial for enabling a community to unify, take action, and sustain this action [66]. In the specific case of LGBTQ+ people, collective identity has been crucial in accomplishing substantive collective action, including resisting discrimination, and organizing for rights and political gain [23, 34]. There is also evidence that a strong collective identity makes internal conflict less destructive and more recoverable [66], while a continued focus on fragmented LGBTQ+ identities seriously limits the opportunity for alliances and solidarity [73].

This tension between group and individual identity is always present when a collective identity is in play [66], but in the case of the overall LGBTQ+ community, this tension is a core conflict that any LGBTQ+ work must grapple with [23, 34, 73, 87]. Practically, there are benefits to both approaches. The queering or "loosening" of LGBTQ+ identity helps deal with cultural sources of oppression – here, cultural ignorance of the variations within LGBTQ+ identity, both in general society and in on-platform communities – while "tightening" around a collective identity helps fight institutional sources of oppression [34]. "Institutional sources of oppression," in the context of this study, include platforms themselves, as they function as a public square with all the attached structural power [35, 36], and have outsized, deterministic impacts on how queer identity is expressed and how queer communities can take shape [1, 41, 42, 53].

Considering our goal of identifying community LGBTQ+ design values for social platforms, we must take seriously the need for solidarity in collective identity when dealing with platforms. However, following Gamson, this collective identity must also have significant internal room made for diversity of identity [34]. Additionally, we take inspiration from prior work on defining and imagining transgender technologies, which have asserted that trans tech (and, accordingly, the values it is based on) must embrace the specific concerns of [1] and what "lie[s] at the heart" of the group being served [41], and be community-focused, accounting for in-group variance [42]. We engaged with a broad sample of the LGBTQ+ community using values elicitation methods which allow the resolution of intracommunity conflict. Our goal was to identify overarching design values for the collectively-identified LGBTQ+ community which can aid in the improvement of current platforms and the creation of future platforms designed to support LGBTQ+ people from the start.

3 METHODS

This study's methods draw inspiration from multiple sources, including VSD, participatory design (PD), and the emerging space of human-centered algorithm design. However, VSD remains our core inspiration, and it motivated our methodological decision-making. In our work, we convened an online group of 31 members of the LGBTQ+ community using a modified version of an asynchronous remote community (ARC) as our deployment framework [61, 62]. In this group, we engaged participants in five distinct activities which adaptively built on each other to move from the promises and pitfalls of design to practical suggestions and considerations of trade-offs in community priorities.

In this section, we will describe our method as motivated by the overall structure for VSD suggested by Friedman et al. [32], as modified by subsequent work (e.g., [6, 56]). It is important to note that we do not claim to have engaged in a full process of value-sensitive design; we have not generated any technical artifact or new system, and rather than engaging the full tripartite

method, we have limited our conceptual work to inductive interpretation of participant values, informed by our positionality as members of the LGBTQ+ community. We have engaged in an empirical value elicitation investigation guided by VSD principles per Le Dantec [56], drawing inspiration for specific activities from participatory design traditions, per Borning and Muller [6]. We deployed this through an ARC approach to elicit relevant design values directly from a group of marginalized participants to support future work, including full implementations of VSD as applied to the LGBTQ+ community.

3.1 Participants

One key element of all VSD is stakeholder analysis, identifying both direct and indirect stakeholders in the context under study [31]. For this study, the research team identified people who participate in online LGBTQ+ spaces as the direct stakeholders, and those members of the LGBTQ+ community who do not commonly participate in these spaces as indirect stakeholders. This, and a desire to recognize the diversity of individual identities and the potentially-differential effects of systems on these diverse identities [17] motivated both our sampling strategy and research framework. In terms of sampling strategy, we needed to include those who are commonly part of LGBTQ+ online spaces, as well as also a group that was reflective of the diversity of the LGBTQ+ populations. We employed Trost's statistically nonrepresentative sampling technique, in which key axes of diversity for the population under study are identified, and sampling is stratified by these characteristics, avoiding many of the biases of a convenience sample [81]. While there are many characteristics which we could have chosen to stratify our sample based on, we chose five sampling characteristics which have previously been demonstrated to have a large impact on one's experience of being LGBTQ+: age [18], gender identity [26], population density of one's childhood and current environment [40], sexual orientation [84], and race [7].

Participants were recruited using Facebook advertisements procedurally targeted towards highly-populated keywords which apply to the LGBTQ+ community, including "gay pride," "genderqueer," "LGBT community," "LGBT culture," "Human Rights Campaign," "transgenderism [2]," "lesbian pride," "LGBT social movements," and "Congressional Black Caucus[3]." Advertisements were limited to US Facebook users. Study advertisements were also distributed via the authors' personal networks, which include online LGBTQ+ groups.

Potential participants were directed to a form where they self-reported the characteristics noted above. A total of 815 potential participants expressed interest in the study. We invited 77 people and 47 people accepted. Over the course of the study, 16 participants dropped out. Ten participants dropped out after the first activity; the other six dropped out later, citing time constraints. This resulted in a sample of 31 participants.

To be admitted to the study group, participants went through our informed consent process. This research, including sampling strategy, early versions of our activities, and overall project motivation, was reviewed by the IRB at the authors' institution and was classified as "not human research" due to the project's focus on improving platform functionality by drawing on community experience, as opposed to documenting and analyzing LGBTQ+ experience specifically. This means the project was not required to maintain direct IRB oversight or follow normal IRB procedures. However, our own ethical stance, which is based in an ethics of care [79, 80], required us to proceed with the same participant protections as a full-review study, including informed consent documentation with full disclosure of participant and researcher rights and

[2] This term is not usually used by transgender people, but was the leading transgender-related keyword available in Facebook's targeting system at the time of the study.

[3] This final keyword was added on the advice of a Black colleague to boost recruiting in this community, which is frequently underrepresented.

responsibilities, expected activities, possible risks, and expected compensation. This stance also informed how we ran the study on a day-to-day basis, as detailed in 3.2.

Due to the project's IRB status, we are restricted from sharing demographic characteristics at an individual level, and report them in aggregate. We achieved a mix of gender identities (26% cisgender male, 16% cisgender female, 29% nonbinary, 16% transgender male, 13% transgender female), sexual orientations (32% gay, 23% bisexual, 16% pansexual, 13% ace spectrum, and 16% otherwise queer), population densities (25% urban, 48% suburban, 23% rural in childhood, 35% urban, 29% suburban, 32% rural now), and ages (range 19-55, M=30.5, SD=10.7). The racial diversity of our sample included 52% white, 13% Black, 6% Hispanic, 10% Asian, 16% mixed, and 3% Native American participants. All participants were active users of social platforms (at least 3x a week), and 58% of participants had experience as moderators or administrators of online groups. This enabled us to account for the not just the needs and wants of average LGBTQ+ group members, but also the practical concerns of the moderators and administrators that would have responsibility for operating the structures proposed here on a day-to-day basis. All participants were based in the United States.

Participants were compensated with $50 for completing at least four of the five activities and could earn an additional $25 for responding to all five prompts and providing meaningful feedback on at least 2 of the responses of other participants. 84% of participants earned the full $75.

3.2 Procedure

We break from Friedman et al.'s original VSD formulation [32], starting with a direct empirical investigation based out of the community under study per Le Dantec et al. [56]. We still proceed with Friedman et al.'s suggested mapping of benefits and harms onto values [32], but do so from an inductive, empirical, participatory stance to center participant voices and concerns per [6, 56].

This inductive stance and our concerns about enabling diverse participation motivated us to employ a modified version of an asynchronous remote community. This entirely-online method centers around a private group on an established social media platform, used as an organizing space to prompt in-community discussion on the topic at hand, engaging participants with not only researcher concerns, but the concerns of their fellow participants [61, 62]. ARC lowers barriers to research participation around travel, available time, potential stigmatization and, by centering our group on a commonly-used platform, technical barriers as well. It has previously been deployed in potentially stigmatizing research contexts such as studies of people living with HIV and rare diseases and in cases where participants are widely distributed [61, 62, 70]. We employed a secret group on Facebook, the most commonly-used social media platform at the time of the study, and participants used the already-familiar Facebook posting, commenting, and reacting affordances to participate in the study.

While past ARCs have largely employed a week-by-week activity structure, our interest in employing methods inspired by participatory and user-centered design involved a need to maintain momentum and in-group interaction in a way that is difficult to maintain over multiple weeks. We compressed the ARC timeline down to 11 days, framed in our recruiting materials as a short online "summer camp" for LGBTQ+ people. Every other day we introduced a new "conversation" for participants to engage in, with the exact nature of conversations 2-5 dictated by the emerging picture of benefits and harms being discussed by participants in previous conversations. For example, while our blue sky exercise was intended to give participants a chance to move beyond the structures of current platforms, participants were clear that their priority was fixing current platforms. Subsequent activities did not attempt to continue to impose an unwanted blue-sky framing. Participation rates per activity were in line with previous ARC studies [60, 61, 62]. A brief overview of each conversation can be found in Table 1, with full details and original text in Supplemental Materials Appendix A.

Table 1. Short Activity Descriptions and Methodological Motivations

Conversation	Motivation	P.R.
Intros and past experiences: Introduce yourself, post a meme that best represents your experiences with online LGBTQ+ spaces, and share your favorite and least favorite things about being in online LGBTQ+ spaces.	Basic icebreaker to further enable co-construction of knowledge via intra-participant discussion and debate [50], and engagement on benefits and harms in order to begin mapping values per [32], as modified for a fully inductive empirical-work-first approach per [6].	100%
Blue sky wants/needs: Think big and lay out your vision of what an ideal online LGBTQ+ space would be, including how people could represent themselves, how people and content can connect, and what rules/standards/modes of enforcement should be present. Engage with at least two other participants' visions.	Unrestricted blue sky exercise to set up context for later prompts, allow the emergence of values and contexts outside the standard set per [6], and combat potential learned neutrality common to marginalized populations by removing reliance on dominant existing structures [67]. Peer feedback requirement begins process of directly interrogating potential values conflicts per [32].	94%
Algorithmic and automatic issues: Select 2 scenarios from a list of 7 likely areas of function for the type of online space you have described in the past two conversations, and tell us how an automated system should be employed/make decisions in these area (including criteria and data needed). *Areas of function: admitting new members, content delivery, content discovery, content filtering, tagging, affinity/subgroup matching, personal matching*	Engages participants directly with key system components in support of better integrating values into the "organizational structure" of the modern platform space, per [32]. Scenario-based approach provides increased saliency to participants [13], situating value elicitation in the participant's own experience [69]. Algorithmic/automatic conversation inspired by emerging work on human-centered algorithm design (e.g., [3, 85]), and prior work showing that scenario-based algorithmic design have been shown to result in algorithms that are more acceptable to users than direct design [57]. Moderation/policy conversation focuses on moderation structure/tools and policy components [52].	94%
Moderation and policy: Select 2 scenarios from a list of 7 likely areas of function for the type of online space you have described in the past two conversations and tell us how the community's moderators and policy should be employed/make decisions in these areas. *Areas of function: extreme content, language & self-expression, community rules, new member admissions, tagging & content warnings, education & onboarding, moderator/admin selection*		87%
Prioritization/role-taking: If we were to launch the platform we've been talking about over the last four conversations tomorrow, what would it look like, how would it integrate into your life/needs, and what role do you see yourself playing on this new platform?	Final reflection on issues discussed during the study, with a specific eye towards further investigation of how the emergent values from prior conversations can be practically integrated into design as well as the individual lives/use cases of participants per [6, 32]. Additional investigation of value conflicts per [32], especially around issues of labor and compensation necessary to enact participant values and technical/structural suggestions.	77%

Note: Righthand column indicates overall participation rate per activity. For full details on each activity including original prompt language, see Supplemental Materials Appendix A.

Our care ethics-based stance led us to organize the online space and deploy our team in a way which focused on attentiveness, responsibility, and responsiveness towards participants [79, 80], including both anticipatory standard-setting and on-the-ground procedure. Study staff, including the authors and three research assistants, actively monitored the group for the duration of the study to prevent and, if needed, mitigate harm, using a schedule and distinct moderation roles, tasks, and procedures for both regular interaction and emergencies. Participants also agreed to follow a code of conduct for the group which was designed to keep conversation within acceptable boundaries (e.g., no hate speech, no threats) while turning disagreement and conflict productive wherever possible. The full code of conduct and moderation policy are available in Supplemental Materials Appendix B. During data collection, study staff ensured compliance with this code while also following up on participant responses, eliciting more detail where appropriate. Staff also encouraged participants to interact with and comment on each other's ideas and concerns using a framework of "yes, and" or "no, but," which elicited value conflicts between participants, a crucial element of a VSD-inspired approach [32].

3.3 Analysis

We employed thematic analysis to analyze our data, using Braun and Clarke's widely-adopted guidance from 2006 [9]. During each conversation, study staff actively worked to familiarize themselves with the data as it became available, keeping running logs of developing themes and initial ideas to aid in later analysis [9]. After each conversation, a research assistant copied all responses out of the Facebook group and created anonymized databook spreadsheets. These logs and databooks were used by the research team for all subsequent analysis.

Moving into the second phase of analysis, all authors independently open-coded the databooks, using a data-driven approach to initial coding of the entire dataset into broad, sometimes overlapping classifications [9]. The first author did so using the MaxQDA qualitative analysis software, while the second and third authors used a combination of the Excel databooks and analog tools such as sticky notes. In phase three, all three authors independently worked to group their codes into larger themes via a process of sorting, combining, and re-verifying codes, using visual representations as an aid [9]. These themes covered several areas of LGBTQ+ experience, and ultimately included the building blocks of our eventual design values.

In phase four, theme review, all three authors reviewed their own themes at the code level for coherency, and then worked together in order to refine our set of candidate themes, working across the entire dataset to ensure their fit to the data as a whole [9]. For the second part of this phase, all authors gathered for an intensive three-day thematic comparison and review process. This process used the first author's coding and themes, informed by the coding and themes identified by the research assistants, as a starting point. Themes were then debated, adjusted, and repeatedly checked against the databooks, with additional coding and re-coding as needed [9]. Themes which indicated values were particularly closely examined, with several candidate themes being compressed down to four potential design values: autonomy, control, representativeness, and inclusiveness. In this phase, we were specifically attentive to the biases each individual author might bring to their own coding, taking advantage of the different positionality each author has regarding the LGBTQ+ community during the comparison process.

In phase five, the authors worked together to formally define the candidate themes and work through a narrative for each based on the related coded excerpts [9], connecting the themes to relevant literature whenever possible. In this process, as is common [9], we discovered that our four candidate value themes were actually best represented as two value themes with two sub-themes each, with self-determination ultimately encompassing autonomy and control, and inclusion encompassing representativeness and inclusiveness.

3.3.1 Position Statement

The research team are all member-researchers in the LGBTQ+ community, including team members who variously identify as bisexual, lesbian, cisgender, and transgender/nonbinary. All study team members participate in various LGBTQ+ online community spaces, and one author also has moderation experience. One study team member is Hispanic, and the remainder of the team is white.

3.4 Limitations

This study has distinct limitations that are important to consider when interpreting our findings. First, like any study that uses a participatory-type method where participants interact, there is the possibility of interference from a social desirability bias. While our participants appeared to be candid and demonstrated a willingness to openly disagree with the research team as well as each other, this potential bias cannot be definitively ruled out.

Second, while this study was not intended to produce results specific only to Facebook, our reliance on the platform for both recruitment and deployment may have limiting effects. When designing the study, we had to balance these potential limitations against our commitment to drawing a broad LGBTQ+ sample while reducing barriers to participation. Facebook's status as the most widely-adopted social platform in the world made it an appropriate choice for including the most participants and avoiding a potentially-biasing technology learning curve. Similarly, using Facebook's wide-reaching, procedurally-targeted advertising system for recruitment allowed us to reach potential participants far outside of our own networks or the online LGBTQ+ spaces this research team is aware of, which are necessarily shaped by our own specific identities. It also allowed us, via the procedural targeting, to include LGBTQ+ people who do not actively participate in online LGBTQ+ spaces or very publicly display their identity. These are crucial respondents considering both the VSD imperative to consider indirect stakeholders [31, 32] and our goal of broadly-applicable LGBTQ+ design values, and we would likely miss these respondents via simple keyword targeting or posting advertisements to a limited subset of LGBTQ+ groups. However, all of this limited us to those participants who either had or were willing to create a temporary Facebook account. It is possible that this, plus locating the bulk of the study on Facebook itself, anchored participants in the model of Facebook when thinking through design decisions, even though our prompts were not platform-specific. Though the design values we identify below appear to be broadly applicable to social platforms with group functionalities, we urge caution in applying our findings to platforms without group functionality. We also call for further work of this type that uses a different platform for deployment and/or focuses on platforms without group functionality.

Finally, this study was specific to a United States-based LGBTQ+ population. While we believe that the methods used here and the overall value disconnect are largely transferrable to other marginalized groups, further work situated within these groups is needed. This may even be true for LGBTQ+ populations in different areas of the world, as our findings are necessarily shaped by the lived experiences of our US-based sample, and therefore by the practical and policy environment of the United States. It is possible that general LGBTQ+ design values look different in different national contexts – e.g., in a context where even being LGBTQ+ is illegal, inclusion in the risk management sense would likely balance safety and the needs of new members in a more safety-dominant way. It is likely that each marginalized community has its own local values, and to better serve these communities, designers should seek out ways of understanding these values. Our approach represents one way of doing so. While our study looks at the anxieties and values of a general US LGBTQ+ participant pool, future studies should investigate the perspectives of subgroups within this population, especially racial and ethnic groups, which may surface different anxieties about social platform use [39], as well as international contexts.

4 FINDINGS

During our data collection period, two contrasting factors became apparent. The majority of our participants confirmed that online LGBTQ+ spaces were crucial to their individual development and social support. However, our participants unanimously expressed anxiety[4] over the current state and future direction of these online spaces. Without being prompted to speak from an exclusively negative standpoint about the platform itself or start from problems requiring fixes, they expressed concerns about the viability of minimizing potential harm from outsiders, from group members, and from platform structures such as content flagging algorithms and moderator toolsets. Ultimately, our participants expressed what they needed from platforms using the language of anxieties. We use the term "anxieties" because our participants expressed their concerns in terms of worry or unease, uncertain that things would improve in the future. In fact, participants often expressed that any move by the platform would ultimately make things worse than they already were. Through our analysis, we found that participants' anxieties were ultimately expressions of values.

By using anxiety as an analytical lens, we see how people's experiences of current social platforms do not fulfill the specific needs of LGBTQ+ users or align with their values. The majority of our participants' anxieties stemmed from a perceived disconnect between their own values (both individual and communal), and the values enacted by the platform through affordances. Better understanding the tension between users' values and the platform's values can help us to design solutions that more closely align these two sets of values. Understanding what prompts user anxieties in their online spaces provides a framework for design work with this population.

It is important to note that many of our participants' examples are framed in the context of existing platforms. As noted in the Methods section, we did initially engage these participants in blue-sky envisioning work designed to move past the restrictions of current platforms. However, the vast majority of participants were clear that they wanted to prioritize fixing existing spaces, which they have invested significant time, effort, and social capital into, rather than designing new platforms. To reflect the values and desires of our participants, we focus on repairing and improving current platforms in these results. However, as Hardy and Vargas identified, there is a younger, well-educated contingent within the larger LGBTQ+ community which is focused on building new systems [47], and these values, while operationalized on current platforms, are broad enough to help guide future, novel platform design work.

In this section, we describe two expressed values that our participants thought were essential for improving online spaces for LGBTQ+ people: *self-determination* and *inclusion*. For each value, we describe our participants' anxieties, how these anxieties translate to a design value, and what design solutions may serve to better enact these values in online spaces. We address the labor implications of these design solutions in section 5.2.2.

4.1 Self-Determination

The first value we identified is **self-determination**, which we define as the ability of an individual and/or group to make decisions about the people, norms, and technical structures they will be impacted by. Our participants were anxious that decisions regarding LGBTQ+ online communities were being taken out of the hands of community members, or were otherwise being made without the specific context of the community taken into account, defaulting to the general status quo of a platform. These concerns applied at two levels: first, to individuals' ability to opt in and out of systems and interactions; second, to who gets to have control over and define criteria for these systems. People concerned about self-determination ask: *Do I have the information and*

[4] Here, we do not refer to anxiety in the diagnostic sense, but rather a general sense of worry or "uneasy concern" about a situation with uncertain outcomes [21].

agency necessary to opt in or out of systems in accordance with my best interests? Are the people or systems acting upon me and my group doing so in a way that reflects what we would choose for ourselves?

4.1.1 Self-Determination as Opt-in/Opt-out

Participants were anxious about their perceived inability to determine what interactions they would encounter on a platform and when they would be acted upon by computational agents. Here, we see a mismatch between what the platform is perceived as valuing (rolling out universal mechanisms for content delivery and interaction) and what LGBTQ+ people value (the ability to determine for oneself if one wishes to be involved with these mechanisms). We also see an embedded clash between what users think computational systems see as "value," e.g. broad content distribution and maximum engagement, and what the user values, e.g. privacy and control over one's own information.

For some people, this can manifest as anxiety over an inability to opt out of involvement with computational systems in the first place. For example, P46 expressed concern over their content being spread by an algorithmic curation system calibrated to garner more engagement when their goal was social support from their existing community:

> Personal posts on Tumblr get picked up by their sharey algorithm (presumably bc a lot of people were commenting or liking to express sympathy), which for me would be a HUGE "burn the site down delete the account change my name move to Mars" situation.

For P46, the fact that there is no opt-out from algorithmic curation was an anxiety-inducing risk. This example illustrates how the values enacted by the platform (e.g., more engagement on content) were mismatched with users' values (e.g., seeking support from a specific community). While social platforms allow users to tweak their privacy settings to only show content to certain people, all but Facebook do not allow users to opt out of algorithmic boosting of their content at a fine-grained level. For instance, Twitter does allow people to opt out of making their tweets public, however this option applies to all of an accounts' tweets (rather than different settings for individual tweets), cutting the user off from the wider Twitter network and the social, informational, and identity support it may provide.

In addition to concerns around algorithmic actors, participants also expressed anxiety about being able to opt in or out of interactions with other platform users. If they did not have the necessary tools to opt out, they were worried they might be unwittingly forced into interactions they had no desire to participate in. For example, P40 noted problems they had encountered in spaces that did not give them control over whether their profile was searchable on the site, or whether they could be directly contacted by strangers:

> Consent is important and people need to be able to say no and expect it to be respected. Otherwise it's not a safe space. One popular BDSM community site went to hell in a handbasket and a lot of people bailed on it when they allowed people to search profiles for age, sex, orientation and location. There was no way to opt out of that search and it raised the already ridiculous amount of random, creepily depersonalized, fetishistic solicitations to an intolerable level. If you do institute that feature, it absolutely has to be opt-in so only people who want to be randomly searched for by people looking for hookups can be.

Because they could not opt out of these features, they were exposed to graphically sexual messages they had no desire to receive. As a result, unwanted interactions with "creepazoids" was a constant source of anxiety in these spaces for P40. Their desire to limit their interaction with strangers was in tension with the platform's seeming desire to introduce them to new people.

4.1.2 Self-Determination as Local Control

Another type of anxiety centered around who determines the criteria on which computational systems base their decisions, and who has the final say over major decisions such as what content should be promoted or flagged and deleted. While the platform may base these decisions on what will maximize platform-wide engagement and minimize the platform's own risk, users wanted these decisions to be based on what is best for their local context. Users worried that because the people or systems making these determinations did not have the appropriate context, decisions were made that were at best misguided and at worst harmful.

One major complaint from participants is that algorithmic systems often filter content in ways that do not align with users' values and concerns. Several participants noted the inability of filters to understand differences in language usage (e.g. LGBTQ+ people using "queer" as a label vs. hostile outsiders using it as a slur[5]). These posts would be inappropriately flagged due to a lack of local context, restricting the community's modes of expression. This disconnect between platform and community standards is more noticeable when compared with how non-LGBTQ+ content is treated, as P10 noted:

> I don't have a lot of trust in algorithms considering how it works on YouTube. Often, purely innocent posts (particularly coming out videos, posts about LGBT+ health, etc) can get flagged by YouTube's algorithm, thinking that it could be hateful content, while actual hateful content might be able to avoid detection from the algorithm by using more coded… language.

For P10, a flagging algorithm that targets LGBTQ+-related content, but not "actual hateful content", indicates that the control over what gets labeled as "hateful" resides with people who do not have the interests of LGBTQ+ users in mind. This experience and distrust of these systems, and the resultant anxiety over how these systems will act on one's own posts, are typical in our data.

Participants were also concerned about their lack of control over content delivery algorithms. Algorithms that are usually configured to boost overall engagement on content may not serve the specific needs of the LGBTQ+ community. As P21 noted, what is engaging to a broad audience may not align with that is engaging with a specifically LGBTQ+ audience. As P45, P46, P27, and P23 discussed among themselves, highly-engaging content may also be harmful. Posts containing hate speech, for example, may garner many reactions and comments, but that does not mean that a person needs or wants to see hate speech at the top of their feed. Moreover, as P28 explained, a feed driven by popularity could both misrepresent the LGBTQ+ community and crowd out new voices:

> I think popularity-based algorithms would better serve some types of content more than others… let's say we're comparing two different posts. One is a selfie of a conventionally attractive, muscular cis gay man who is a long-time user with a lot of followers. The other is a text post from a newer user with less followers who is venting about a bad day dealing with microaggressions at work and feelings of dysphoria. Because the selfie has high engagement, it is brought onto more people's feeds and continuously gets higher engagement, while the other post is consistently buried low on everyone's feeds. I think these popularity-based algorithms can hinder efforts toward inclusivity by consistently highlighting only certain types of narratives, which feels counter-intuitive to the goals of an LGBTQ+-focused site.

The majority of our participants highly valued keeping control over the community within the community as much as possible, and saw automatic platform decision-making structures largely as unwanted, decontextualized algorithmic interlopers.

[5] The status of the term "queer" is not fully agreed upon within the LGBTQ+ community. While many view the term as a reclaimed label for the overall community, others still view it primarily as a slur [34].

4.1.3 Enacting Self-Determination Through Design

We now turn to design solutions our participants suggested, which may serve to better enact the value of self-determination in online spaces. These suggestions centered around providing clearer expectation-setting for users and moving the locus of control over an online space from platform-wide algorithms back to individual users and communities.

Our participants suggested that online spaces could enact better expectation-setting at multiple levels. For the initial decision to opt into a space, many participants noted a need for clear community and platform rules. These rules should clearly serve as expectation-setting and scaffolding for new members, and should include mechanisms for ensuring that new members have in fact read and understood what exactly they are agreeing to participate in, and what mechanisms will be acting upon them. P18 noted that one such mechanism is often informally employed by Facebook group moderators, where new members are quizzed about the group's code of conduct when requesting to join a group.

In terms of opting in to encounter content in the first place, participants suggested formalizing and expanding the prominence and importance of tagging systems. Several participants suggested making content tags mandatory for one's post to be included in algorithmic content distribution. Doing so would ensure that users have clear expectations about the content they engage with, as well as the tools to avoid certain types of content if they wish. Multiple participants cited Archive of Our Own (AO3) as an example of a platform which handles these kinds of issues well, as users who wish to post a story must use tags to signal whether their story may contain upsetting content such as descriptions of violence, thus providing other users with the necessary information to avoid violent stories . By giving people the information and tools necessary to opt in and out of certain spaces, and certain experiences and content in these spaces, these design solutions support self-determination.

Our participants also advocated for an increase of local context in automatic systems to support ongoing self-determination. Some participants suggested that the parameters and criteria of automatic systems be set locally, at the level of a community (e.g., an online group), rather than at the level of the platform overall. For example, P39 suggested that if a word-level content filter is indeed necessary, then the list of banned or problematic words should be sourced from the LGBTQ+ group itself, reflecting community norms:

> …on the topic of the banned word list, I'm assuming that the list would be curated based on a small group or forum within the main site and not something site-wide. If a content filter is absolutely necessary, then it should be specific to the group's needs

By rooting filtering criteria in the group's context and basing them on the group's values, a platform may avoid misunderstanding and losing user trust down the road.

Similarly, participants advocated for feed algorithms to consider factors other than overall engagement, instead prioritizing content based on more germane community standards. For example, P28 and P46 suggested curation based on the content itself, taking the example of a group dedicated to social support which would prioritize question-asking over selfies. P23 and P46 suggested that an algorithm prioritize user characteristics such as new member status and prior lack of engagement, if one of the group's goals is to boost inclusivity by encouraging new and lurking users to participate. On an individual level, participants suggested that algorithms screen content in or out of a user's feed according to the user's preferences and values, for instance by prioritizing certain content tags and excluding others.

Many participants suggested that users should maintain ultimate control within their group environment, with all decisions made by an algorithmic system subject to review, and always the option for users to appeal to a human administrator (preferably a fellow community member, rather than a platform representative). As P20 explained, while there are cases where algorithmic review makes sense, the overall desire is for a hands-off system:

> …unless there is like nudity involved or like a blatant violation, then okay system, do your thing, but if otherwise it should have to be confirmed by human eyes

Several participants suggested this could be accompanied by new moderation tools to enhance human decision making, while also providing the necessary context for why an algorithm has suggested a particular course of action. For example, when an algorithm flags a user's post for deletion and/or a ban, a moderator would receive useful statistics on the user's past behavior alongside the offending post. This would allow the human moderator to understand why the algorithm has identified this user as a bad actor, and make their decision based on not only the information provided by the algorithm, but also the community's values.

Concerns around self-determination are largely driven by a disconnect between what users and communities wanted and what the platform chose for them. Our participants pointed out that self-determination should be respected by a system over time. Self-determination is not a one-time concern, and a group's local context is not suddenly invalid because a system introduces an upgrade or new feature. By giving communities and individuals the information and tools to act according to their best interests, and not acting upon them without their approval, platforms can support people's self-determination.

4.2 Inclusion

The second value we identified based on participant responses is **inclusion**, which we define as the ability of an online group to widely welcome, support, and reflect the breadth of the community the group engages with while still maintaining group safety. This tension mirrors the larger tension in the LGBTQ+ community between solidarity and diversity, but at the level of group dynamics rather than population scale. Participants expressed anxieties around the vulnerability of individuals in LGBTQ+ spaces to harms originating from both outside and inside the community [74]. However, they also described concerns over harming others by excluding LGBTQ+ people who may need such spaces, putting broad inclusion directly in tension with group safety. Within groups, anxieties about inclusion take the form of concerns about how reflective group governance of different parts of the community, and how responsive group rules, standards, and norms are to change over time. People concerned about inclusion ask: *Do I and others feel safe and welcomed in this space? Do our administrators, moderators, and policies reflect the composition and values of the group as it stands today?*

4.2.1 Inclusion in Risk Management

Anxieties over risk management concern the permeable boundaries of online spaces. On one hand, participants reported anxieties around the potential for harm from bad actors. P37, for example, described negative experiences in spaces with open membership policies:

> My perspective is colored by having been a tumblr user when bigots from another site would plan 'raids' and fill the queer hashtags with violent images and hate speech. If anybody can sign up at any time from an open site, it poses a serious danger to the platform's users.

While such concerns over outsiders entering spaces disruptively were most common, a subset of our participants were also concerned about the potential for bad actors to enter LGBTQ+ spaces to gather and repost content outside the group for mocking or outing individuals, as P10 noted:

> …even if it's a non-anonymous space with people I don't know in real life, I'd rather not have the opportunity for someone to spitefully blackmail me or something by showing my grandmother pics of hot anime dudes I posted somewhere.

This anxiety over harm from these bad actors created a strong incentive to screen potential members.

On the other hand, this protective instinct was in tension with the desire to be as welcoming as possible. For example, P23 described how useful they found it to have conversations with

people who represent a range of ages, or whose identities are less understood in LGBTQ+ spaces, such as aromantic and intersex people. Several participants discussed the importance of welcoming people who are exploring their identity, as well as those whose identity is more solidified. For participants, this value of inclusion often manifested as an anxiety over accidentally excluding those who in fact should be welcomed in the space.

Tensions over risk and inclusion also emerged in discussions around language norms. Participants noted how some members of LGBTQ+ spaces are uncomfortable with specific terms, such as the word "queer". In some communities, to protect users from hurtful language, posts containing the word "queer" may be automatically deleted, and users who repeatedly use the term may be automatically banned. However, as P21 noted, these reactions may be exclusionary to those that are still learning:

> I see a lot of conversations being shut down [over language], and I perceive it as being unnecessarily exclusionary. I think that the internet is here to be used for learning. And I personally have a hard time tolerating things I see as exclusionary, or judgmental of people who are simply less educated at this time.

Several participants point out that some discomfort, or acceptance of risk, can be productive in discussions of complex issues. For instance, P5 noted how important it is to talk about racial issues in LGBTQ+ spaces, even though some might be uncomfortable with the topic:

> I think that most people don't like how uncomfortable talking about racism and how it [racism] affects all of us – whether the effects are positive or negative – feels like. I think that being uncomfortable is an essential part of those conversations, because thinking about what privileges or disadvantages you have is definitely something that not everyone does.

Automatic censoring and bans may discourage newcomers and curb important discussions. However, unfettered exposure to certain topics or language may make group members feel unwelcome and unsafe. Finding a balance between protecting community members from harm and encouraging a community that is welcoming and encourages thoughtful discussion of sensitive topics is a challenge for LGBTQ+ online group spaces.

4.2.2 Inclusion in Governance

Our participants also expressed anxieties around the roles of group moderators and administrators, specifically who holds power within a group and whether policy and moderation structures reflect overall group composition and values. As a community's composition and values evolve, participants expressed concern over whether the power structure of a community is able to evolve in kind. This desire for a sustained, internal inclusiveness is in tension with current platform tools, which often let group founders retain near-absolute power indefinitely.

Participants expressed anxiety about how community members with authority (e.g., moderators and administrators) came to positions of power and who had the power to write community rules. On many platforms, the founders of a group are given that authority - someone who creates a Facebook group can dictate the group's privacy settings and code of conduct. However, as many participants pointed out, there is no guarantee that the founders of a space are representative of the current membership and inclusive of all subgroups, a major concern of many participants. This in turn raises anxieties over how community members who are not included in leadership will be treated by others. P6 shared concerns over what they saw as a misalignment between moderators and the community as a whole:

> I've noticed that in a lot of the communities I've been in, somehow the mods (by some weird coincidence) seem to often be white, lesbian trans women, and I think that a lot of times people from other groups are underrepresented, especially trans men of color. I'd love to see a spectrum of both identity and background, because not everyone has the same experiences.

P6 noted that this misalignment could lead to false consensus over what is and is not hurtful to community members. P43 suggested that misalignments between community composition and leadership could have an impact on policy enforcement as well:

> ...the mod teams [on a site] ought to be heavily representative of QPOC [queer people of color] and disabled folx. Anti-bigotry rules can only go so far if they're enforced by a group with no direct experience with the subjects the rules apply to.

Even if the policies of an online space are intended to protect certain group members, their effectiveness will be limited if they are enforced by members who may not be sensitized to identify the relevant kinds of interactions these policies are supposed to help protect users from. Having moderators that are representative of the group is crucial for addressing the needs of all community members.

This recognition that a community will evolve, and the anxiety that the composition and values of those in authority may not evolve with it, point to concerns around growth and flexibility in general. Participants recognized that the dynamics and moderation needs of a group at its founding may be very different than that same group a year later, particularly if that group has grown. As P23 noted, "*a good and safe online community needs to be really well-moderated as soon as it isn't really small.*" What "really well-moderated" means, however, varies based on the values of the community in question.

4.2.3 Enacting Inclusion Through Design

Participants in this study proposed many solutions for protecting their communities against bad actors while maximizing inclusivity. Suggestions for how to manage the general tension between safety and welcome fell into three categories: education, sandboxing, and multi-tiered group structures.

First, participants suggested tools to encourage education, especially around language and norms. Gentle interventions could help newcomers learn the ropes without making new community members feel unwelcome. For instance, P36 suggested a system that would automatically explain why language might be unwelcome:

> Maybe some kind of little dictionary of identity words that are considered hurtful can pop up when a word is detected and can be accessed from an education section of the website? It could be gentle, "hey bud, this might not be appropriate," without jumping down throats.

By employing the computational system to educate rather than punish group members who use hurtful language, platforms can encourage a balance between inclusion for good-faith newcomers and safety from harmful behaviors.

Second, our participants had suggestions for sandboxing new members, or initially limiting their access, privileges, and/or abilities. Several participants suggested that an online LGBTQ+ space might consider having a broad admission policy, with the caveat that new members would only have restricted access to the group. Some participants suggested that new members be restricted to only seeing a subset of "low-risk" content, such as posts the group had marked as "less sensitive" and/or unlikely to be harmful if shared outside the group. Only once the new member had passed a set probationary period, and/or a human moderator approved full membership, could they join the full group. Other participants suggested systems that could progressively unlock access levels. For instance, once a new member had regularly commented without requiring moderator intervention, or after a certain number of established group members had vouched for them, the new member could unlock full posting permissions. These measures may maximize the inclusivity of online spaces while minimizing the potential for harm.

Third, participants suggested implementing a tiered group structure, featuring an overarching LGBTQ+ community space for building community solidarity and maximum inclusivity. Subspaces (sometimes conceptualized as "rooms" within a larger house), would be organized

around different identities, or different norms of behavior. In discussion, P23 and P47 best expressed the group's overall reasoning for this solution:

> I like the idea of forums and subforums or some kind of analogue for the community, because I enjoy the experience of having a broader group of people that feels safe and open, surrounding smaller, more tightly-knit, more niche spaces. (P23)

> That way folks could have their general, everyone's-welcome discussions and people who need smaller spaces with similar experiences would know where to find them without risking that contempt or unwelcome interaction. (P47)

Tiered group structures are a promising avenue for both inclusion and a hyperlocal form of self-determination. Participants could participate in larger umbrella spaces that allow for solidarity and community building and still choose subspaces as they wished, thereby enacting self-determination. Allowing sub-spaces with their own local norms for discussion would keep the locus of control inside the subcommunity. Finally, moderators and admins of subspaces could be selected for a particular space, enacting a localized version of inclusiveness.

On the level of enacting inclusion in group leadership, the majority of participant suggestions centered around democratizing and distributing in-group authority, and on-platform mechanisms could provide useful structures for doing so. Half of the participants suggested that, rather than group founders holding onto power until they voluntarily abdicate, a regular election mechanism could be implemented for groups. Such a mechanism would open moderator and admin roles to a vote, and also provide an opportunity for the whole community to revise and vote on a new version of the group rules or code of conduct. More generally, platforms could provide structures for groups to self-determine their own governance structure, e.g. option to formally set and have the platform enforce term limits and regular election date if the group selects a democratic structure.

Many participants proposed moderation structures that could distribute moderator powers and involve more members of the community. For instance, P40 and P41, in discussion with several other participants, proposed a "community moderation" setup, where certain moderation functions could be taken over by volunteers within the community. On-platform tools to support these structures could include additional tiers of authority in the moderation setup, granting volunteer moderators some power over certain areas. For instance, a lower-tier moderator could have the power to delete or mute individual comments flagged by other users, but only a full moderator could ban someone from the community, similar to how certain Reddit subreddits currently function. This setup may provide several benefits: decentralizing power in online groups, allowing distributed decision-making on less critical issues, providing a sense of ownership to community members, and acting as training for those interested in becoming full moderators/admins in the future. On-platform tools to support these structures could include adding additional tiers of authority in moderation setup, granting volunteer mods some power over certain areas. For instance, a lower-tier moderator could have the power to delete or mute individual comments flagged by other users, but only a full moderator could ban someone from the community. However, it is important to note that other structures proposed here would have to apply to all tiers of such a hierarchy – top-level administrators in a democratized group, for example, would need to be just as subject to elections as lower-level moderators.

While concerns around self-determination are largely driven by a disconnect between what users and communities wanted and what the platform chose for them, concerns around inclusion instead center around the dynamics and governance of the community itself. However, platforms can still support people's desire for inclusion in online spaces in order to help mitigate some of the anxieties described here. Design solutions that balance safety and broad inclusion, prioritize education over exclusion, and democratize and distribute in-group authority, can all go a long way toward supporting LGBTQ+ communities online.

5 DISCUSSION

By engaging LGBTQ+ participants in a series of value elicitation exercises per [56], we found evidence of a misalignment between what LGBTQ+ users perceive as the current values of social platforms and the values of LGBTQ+ users themselves – a misalignment which was a source of anxiety for our participants. This VSD-inspired approach helped isolate two design values based in the lived experiences of members of this group: ***self-determination*** and ***inclusion***. Prioritizing these values (and the design solutions proposed by our participants) in future design work can help platform designers reduce this misalignment. Considering the role of social platforms as instrumental tools for LGBTQ+ people [5, 16, 18, 25, 29, 38, 45, 49, 59] and the lack of alternatives [22], addressing this misalignment is crucial to better supporting LGBTQ+ users.

The instrumental nature of social platforms for LGBTQ+ users may motivate some of the value disconnect at hand. Whereas platforms [35, 36] as well as prior work in HCI (e.g. [55]) value goals such as community growth, user retention, and increasing engagement, our participants instead valued productive, safe, and inclusive group membership, as opposed to a large or active membership. Platforms' focus on engagement was specifically identified as misaligned, with engagement-based curation often promoting material our participants found harmful or frivolous instead of material which served a purpose for the group, such as requests for validation or information. Our participants found this to be a violation of *self-determination*. Likewise, our participants did not want to retain every user or grow groups indefinitely, nor did they want hyper-focused spaces built around a singular identity. Instead, they valued careful *inclusivity* which recognizes the harm bad actors can do while also acknowledging possible limits to growth for a group with instrumental purposes such as health information distribution and identity formation [16, 18, 29, 49, 59]. This is in contrast to work such as Ren et al.'s, which notes that internal diversity can at times be harmful to a community by damaging user retention; the goal here is not user retention as Ren et al. identify [72], but rather building solidarity within a broad, multi-faceted marginalized population [65]. The spaces our participants want do in fact fulfill member needs as noted above, but participants did not use this as a user retention tool as predicted by Ren at al. [72], highlighting the need to think about groups formed out of shared identities differently than those that build a shared identity as group members.

It is in the best interest of platform operators to identify and address these value disconnects, which pose a direct challenge for their designers. People prefer to engage in spaces that support their identity and make them feel like their "true selves," [75] and a platform with values that conflict with user values makes this difficult to achieve. Past work suggests that value disconnects already motivate LGBTQ+ people specifically to limit their engagement with or leave platforms [11], suggesting that value misalignments not only cause platforms to fail to support users by making it harder to access crucial information and services, but may also harm their own goals of growth and user retention. Prior work on social media ecosystems suggests that LGBTQ+ people already view social platforms at an ecosystem, not a platform, level, and are capable of shifting their attention and effort elsewhere if they must [20]. Platforms could address this challenge by better understanding the diversity of values that exist within the many communities each platform hosts and providing multiple toolsets for building and maintaining communities with divergent contexts.

Overall, we offer these values not as a new universal set of values, but rather a situated set of values for addressing the design challenge around value misalignments noted above and improving the state of online spaces for a marginalized group, the LGBTQ+ community in the US. While we believe these values have relevance in other marginalized communities, future work in the context of other communities is essential, and we suspect localized formulations will differ on some dimensions from what we have found here.

5.1 Designing for the Diversity and Solidarity Challenge: Paired Values in Balance

While these values have the potential to help platforms design for (and, ideally, with) LGBTQ+ people generally, these values also help to address the complex design challenge of respecting the diversity of the LGBTQ+ community while also supporting group identity [34]. Gamson noted this larger tension between a solid group identity and recognition of the internal diversity of identity, that "the challenge...is not to determine which position is accurate, but to cope with the fact that both logics make sense" [34 p. 391]. In the case of addressing this tension via social platform design, the challenge remains the same: not to prioritize one of these two principles, but rather to cope with the fact that they must be held in balance to be functional. In this section we will demonstrate how designing solutions that enact *both* self-determination and inclusion can help navigate the tension between group identity and internal diversity.

Consider the core theme behind the value of *self-determination*: keeping the locus of power close to the community itself. Participants wanted local standards and algorithmic actors specific to and calibrated by their local LGBTQ+ community, and the ability to opt out of large, potentially harmful content distribution structures. Effectively, this value calls on platforms to let online LGBTQ+ spaces have power over themselves. In doing so, this largely fulfills the core promise of building a group identity for LGBTQ+ people in the first place [23, 34, 66, 73]. Systems that respect and enact self-determination would let LGBTQ+ groups have more control over their relationship to the platform's structures, which in turn creates an opportunity to practice internal queering and recognition of diversity without surrendering external organizing power. In other words: enacting self-determination becomes a check against institutional sources of oppression.

With the line held against outside forces via enacted *self-determination*, there is then enough room to successfully enact the value of *inclusion,* which can hold the line against internal and cultural sources of oppression. It is important to note that inclusion as laid out in this study requires us to cope with the dynamics of the internal diversity of the LGBTQ+ community. This process is essential because current structures still act to amplify *intra*community stigmatization and exclusion [74, 83], the exact kind of intracommunity behaviors that make it essential for platforms to recognize and actively support this internal diversity [23, 87]. Participant solutions such as scheduled leadership elections and more varied and accountable moderator roles take a general, somewhat vague commitment to respecting this diversity and make it mandatory via platform structures. In other words, designing for the sort of inclusion we have discussed here can help prevent known intracommunity harms (e.g., [74, 83]) while also pushing back against the general cultural lack of understanding of the internal dynamics of LGBTQ+ diversity.

Without inclusion, even systems that fully enact self-determination could continue to be harmful. A system which allows administrator/moderator review of every algorithmic determination that impacts the group does not guarantee that, for example, a cisgender gay moderator will treat cisgender and transgender content equally during that review process. Similarly, without self-determination, a system that fully enacts inclusion could still allow continued harm. A system which enforces regular leadership elections could result in a moderator team that reflects the diversity of the LGBTQ+ community, but if the majority of decision making is still left to context-free algorithmic systems, this leadership team has no real power. By balancing the two - by making sure, for example, that we have both inclusive moderator teams *and* tools which support self-determination via local control - we can make space for both collective identity for solidarity and internal diversity.

5.1.1 Backwards-Compatible Values: Attending to Marginalized Subcommunities

If a commitment to recognizing internal diversity is a priority, we must also be attendant to the fact that subcommunities under the LGBTQ+ umbrella do sometimes have their own sets of design-relevant concerns, as past work on LGBTQ+ subcommunities has shown (e.g. [1, 42, 47, 74, 83]). Ignoring these concerns would be problematic, but centering design on the two

overarching LGBTQ+ values we have identified here does not require us to abandon more specific subcommunity values, and can in fact directly support the enactment of those value sets. Consider the still internally-stigmatized transgender community [74], for which both Ahmed et al. and Hamison et al. have promulgated value-like design [1, 41, 42]. Much of this work is directly compatible with and supported by self-determination and inclusion.

A focus on self-determination in design, and especially local control of content standards, directly addresses the issue of trans-related content not being compatible with broad "family-friendly" platform standards [41]. With control over content and behavioral norms set locally there would be both the space and the community moderation expertise to safely allow the type of content Hamison et al. refer to as "erotics," which can include crucial medical information and self-exploration content but is often broadly labeled as pornographic by platforms using broad, generic standards [41]. Similarly, local control over standards of what a "real" identity is can combat a harmfully-generic and permanent platform conception of "real" identity is [43]. This supports trans-positive aspects of technology such as a more LGBTQ+-compatible standard for realness and support for identity exploration and change [41]. Additionally, trans-specific security concerns around consent and disclosure [58] would be directly addressed by the opt-out form of self-determination, especially if, as our participants suggested, they may opt out of automatic algorithmic content distribution.

A focus on inclusion is also broadly compatible with prior work on designing for trans populations, which recognize the same tension between inclusive representativeness and security [58] that we see in our participant discussions. This work has found that transgender users of social platforms value diverse, supportive spaces [41]. Enacting inclusion should support this goal through mechanisms like sandboxing which allow for broader admission to groups. Similarly, while the tension between inclusiveness and security will likely always be present, one way to handle it is through governance structures. Ensuring that the leadership which is in charge of managing the tradeoffs reflects the diversity of the group makes it more likely that these decisions, in turn, will support both the separation of content that is crucial to transgender users [41] as well as broad inclusiveness in membership. Additionally, the types of mechanisms our participants have suggested to help newcomers acclimate to local norms, allowing both inclusive discussion as well as behavioral standards, could be helpful in supporting the openness to serious, sometimes sensitive discussion that is a crucial aspect of trans-friendly technology [41]. Simple prompts around language and full post approval only after a sandbox period can help create an atmosphere which avoids intracommunity fights over identity expression [83] even when discussing such charged, emotional topics. Ultimately, by enacting broad-based LGBTQ+ design values such as self-determination and inclusion, we also support the specific needs of key subcommunities.

5.2 New Challenges for Design

In addition to this pair of localized design values, this research has generated practical suggestions to implement these values, sourced from the community itself. To enact *self-determination*, participant suggestions centered around more, and more visible, on-platform expectation setting, expanded tagging systems which can be used as filters to avoid content, and an increased approval-focused role for moderators and administrators on most types of algorithmically-made decisions. To enact *inclusion*, participant suggestions centered around early, potentially-computational interventions to educate new members on norms; sandbox spaces for newcomers; and tiered structures that allow the presence of an overall LGBTQ+ group to build solidarity as well as smaller subordinate spaces which recognize specific identities. We urge designers to examine each value's suggestions section (4.1.3 and 4.2.3, respectively), which discuss these solutions in more detail - if not for direct implementation, to anchor design processes in the needs and system understandings of the community.

In further examining these values, we also identified two cross-cutting issues which represent additional meta challenges to future design work with an LGBTQ+ population. We present these not as design implications, but rather large-scale, ongoing challenges for future design which our participants regularly brought up but found difficult to grapple with. With further study, they may become contextualized values themselves: trust and labor equity.

5.2.1 Trust

Our second conversation was a blue-sky exercise, but the research team grappled with limitations the participants imposed on themselves: a pervasive sense that trust in a computational system is unwise, especially trust in a system's ability to be more responsive to user values. This reflected the distrust in platforms and a seeming inability to limit data collection or control one's information identified by Marwick and Hargittai [63]. This suggests that, even when aiming to improve conditions for marginalized users, designers face an uphill battle to reclaim trust with possible implications for system use. This highlights the importance of considering trust when examining any problem involving a social platform, as humans have expectations around trust that are difficult for systems to satisfy, but which must still be addressed [33]. To improve conditions on social platforms for marginalized groups, designers will have to grapple not only with technical implementation, but ways to explain changes to rebuild bridges with marginalized groups.

5.2.2 Labor Equity

Many approaches our participants favored to enacting their collective values were based in returning control to human users, as opposed to computational systems, or otherwise required human labor from community members to implement. Many of our participants were aware that their suggestions required labor, and repeatedly noted this during discussions, and most participants directly grappled with this need for additional labor during the final discussion around value conflicts and role-taking. However, it was clear that not everyone suggesting such changes was interested in taking up a moderator role themselves. Studies on peer production have long recognized that a small group of participants do the bulk of the labor [76, 78], and over time control becomes further concentrated among elites [76]. This outcome contradicts the inclusion our participants value.

We recognize that this will be a challenge to address, but suggest that attending to the values noted here would be a good first step. By way of example, attending to the values of the fan fiction community, defined in opposition to older platforms where these values were ignored or violated, was crucial to the success of Archive Of Our Own [27]. That platforms adopted a strategy of engaging community members in peripheral participation to build volunteer capacity which has been effective in helping to sustain the community [28]. Designers could explore community-based moderation setups like the one suggested by our participants to distribute the increased labor, or institute formal mechanisms for rotation. Of course, issues around the ethics of involving humans in moderating potentially-traumatic content will still need to be addressed [36].

6 CONCLUSION

In this study, we have engaged with a diverse group of US LGBTQ+ participants, identifying perceived values mismatches with social platforms which currently cause LGBTQ+ users anxiety. As social platforms continue to play key instrumental roles for LGBTQ+ users [16, 18, 25, 29, 45, 48, 49, 86], we hope the community-sourced values we have presented here, along with our participant's suggestions on how to enact these values, aid social platforms in developing tools to better support these marginalized users while preventing further harm. Overall, social platforms have enormous opportunity to support marginalized communities – as our study

suggests, the key to realizing this potential may be engaging with these communities directly in order to honor and better align with their local values.

ACKNOWLEDGMENTS

We acknowledge the crucial work of our research team: Sam Hassett, Emily Hoecker, and Brianna Dym. We acknowledge the time and effort of our participants, and thank them for sharing their experiences and participating in discussions. Finally, we thank Stevie Chancellor, Emily Wang and Mark Diaz, as well as the reviewers and associate chair, for their helpful feedback.

REFERENCES

1. Alex A Ahmed. 2018. Trans Competent Interaction Design: A Qualitative Study on Voice, Identity, and Technology. *Interacting with Computers*, 30, 1: 53-71.
2. Nazanin Andalibi. 2019. What Happens After Disclosing Stigmatized Experiences on Identified Social Media: Individual, Dyadic, and Social/Network Outcomes. In *Proceedings of 2019 CHI Conference on Human Factors in Computing Systems*, 137.
3. Eric PS Baumer. 2017. Toward human-centered algorithm design. *Big Data & Society*, 4, 2: 1-12.
4. Greg Bensinger and Reed Albergotti. 2019. YouTube discriminates against LGBT content by unfairly culling it, suit alleges. Retrieved from https://www.washingtonpost.com/technology/2019/08/14/youtube-discriminates-against-lgbt-content-by-unfairly-culling-it-suit-alleges/
5. Lindsay Blackwell, Jean Hardy, Tawfiq Ammari, Tiffany Veinot, Cliff Lampe, and Sarita Schoenebeck. 2016. LGBT Parents and Social Media: Advocacy, Privacy, and Disclosure During Shifting Social Movements. In *Proceedings of 2016 CHI Conference on Human Factors in Computing Systems*, 610-622.
6. Alan Borning and Michael Muller. 2012. Next steps for value sensitive design. In *Proceedings of SIGCHI Conference on Human Factors in Computing Systems*, 1125-1134.
7. Lisa Bowleg. 2013. "Once you've blended the cake, you can't take the parts back to the main ingredients": Black gay and bisexual men's descriptions and experiences of intersectionality. *Sex Roles*, 68, 11-12: 754-767.
8. Judith B Bradford. 2008. Demography and the LGBT Population: What We Know, Don't. *The Fenway guide to lesbian, gay, bisexual, and transgender health*25.
9. Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology*, 3, 2: 77-101.
10. Pam Briggs and Lisa Thomas. 2015. An inclusive, value sensitive design perspective on future identity technologies. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 22, 5: 23.
11. Jed R Brubaker, Mike Ananny, and Kate Crawford. 2016. Departing glances: A sociotechnical account of 'leaving' Grindr. *New Media & Society*, 18, 3: 373-390.
12. Matthew Carrasco and Andruid Kerne. 2018. Queer Visibility: Supporting LGBT+ Selective Visibility on Social Media. In *Proceedings of 2018 CHI Conference on Human Factors in Computing Systems*, 250.
13. John M Carrol. 1999. Five reasons for scenario-based design. In *Proceedings of 32nd Annual Hawaii International Conference on Systems Sciences*, 11.
14. Alexander Cho. 2018. Default publicness: Queer youth of color, social media, and being outed by the machine. *New Media & Society*, 20, 9: 3183-3200.
15. Niraj Chokshi. 2017. YouTube Filtering Draws Ire of Gay and Transgender Creators. Retrieved from https://www.nytimes.com/2017/03/20/technology/youtube-lgbt-videos.html
16. Shelley L Craig and Lauren McInroy. 2014. You can form a part of yourself online: The influence of new media on identity development and coming out for LGBTQ youth. *Journal of Gay & Lesbian Mental Health*, 18, 1: 95-109.
17. Kimberle Crenshaw. 1990. Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stan. L. Rev.*, 43, 1241.

18. Andrea Daley, Judith A. MacDonnell, Shari Brotman, Melissa St Pierre, Jane Aronson, and Loralee Gillis. 2017. Providing health and social services to older LGBT adults. *Annual Review of Gerontology & Geriatrics*, 37, 143.
19. Xuefei Deng, KD Joshi, and Robert D Galliers. 2016. The duality of empowerment and marginalization in microtask crowdsourcing: Giving voice to the less powerful through value sensitive design. *Mis Quarterly*, 40, 2: 279-302.
20. Michael A. DeVito, Ashley M. Walker, and Jeremy Birnholtz. 2018. "Too Gay for Facebook:" Presenting LGBTQ+ Identity Throughout the Personal Social Media Ecosystem. *Proceedings of the ACM on Human-Computer Interaction*, 2, CSCW: 44.
21. Oxford English Dictionary. *"anxiety, n."*. Oxford University Press
22. Petra L. Doan and Harrison Higgins. 2011. The demise of queer space? Resurgent gentrification and the assimilation of LGBT neighborhoods. *Journal of Planning Education and Research*, 31, 1: 6-25.
23. Peter Drucker. 2011. The fracturing of LGBT identities under neoliberal capitalism. *Historical Materialism*, 19, 4: 3-32.
24. Stefanie Duguay. 2019. "There's no one new around you": Queer Women's Experiences of Scarcity in Geospatial Partner-Seeking on Tinder. In *The Geographies of Digital Sexuality*, (ed.). Springer, 93-114.
25. Brianna Dym, Jed R Brubaker, Casey Fiesler, and Bryan Semaan. 2019. "Coming Out Okay": Community Narratives for LGBTQ Recovery Work. *Proceedings of the ACM on Human-Computer Interaction*, 3, CSCW: 154.
26. Laura Boyd Farmer and Rebekah Byrd. 2015. Genderism in the LGBTQQIA community: An interpretative phenomenological analysis. *Journal of LGBT Issues in Counseling*, 9, 4: 288-310.
27. Casey Fiesler, Shannon Morrison, and Amy S Bruckman. 2016. An archive of their own: a case study of feminist HCI and values in design. In *Proceedings of 2016 CHI Conference on Human Factors in Computing Systems*, 2574-2585.
28. Casey Fiesler, Shannon Morrison, R Benjamin Shapiro, and Amy S Bruckman. 2017. Growing their own: Legitimate peripheral participation for computational learning in an online fandom community. In *Proceedings of 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 1375-1386.
29. Jesse Fox and Rachel Ralston. 2016. Queer identity online: Informal learning and teaching experiences of LGBTQ individuals on social media. *Computers in Human Behavior*, 65, 635-642.
30. Christopher Frauenberger, Katta Spiel, Laura Scheepmaker, and Irene Posch. 2019. Nurturing Constructive Disagreement-Agonistic Design with Neurodiverse Children. In *Proceedings of 2019 CHI Conference on Human Factors in Computing Systems*, 271.
31. Batya Friedman, Peter H Kahn Jr, and Alan Borning. 2002. *Value Sensitive Design: Theory and Methods*. University of Washington Technical Report 02-12-01
32. Batya Friedman, Peter H Kahn, and Alan Borning. 2006. Value Sensitive Design and Information Systems. In *Human-Computer Interaction and Management Information Systems: Foundations*, Ping Zhang and Dennis Galletta (ed.). M.E. Sharpe, Armonk, NY, 69-101.
33. Katie Z Gach and Jed R Brubaker. 2019. Experiences of Trust in Post-mortem Profile Management. *ACM Transactions on Social Computing*, ahead of print.
34. Joshua Gamson. 1995. Must identity movements self-destruct? A queer dilemma. *Social problems*, 42, 3: 390-407.
35. Tarleton Gillespie. 2010. The politics of 'platforms'. *New media & society*, 12, 3: 347-364.
36. Tarleton Gillespie. 2018. *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press, New Haven, CT.
37. GLSEN, CiPHR, and CCRC. 2013. Out online: The experiences of lesbian, gay, bisexual and transgender youth on the Internet. Retrieved from https://www.glsen.org/article/experiences-lgbtq-youth-online
38. Amy L Gonzales. 2017. Disadvantaged minorities' use of the Internet to expand their social networks. *Communication Research*, 44, 4: 467-486.
39. Kishonna L Gray. 2012. Intersecting oppressions and online communities: Examining the experiences of women of color in Xbox Live. *Information, Communication & Society*, 15, 3: 411-428.

40. Mary L Gray. 2009. *Out in the country: Youth, media, and queer visibility in rural America*. NYU Press, New York, NY.
41. Oliver L Haimson, Avery Dame-Griff, Elias Capello, and Zahari Richter. 2019. Tumblr was a trans technology: the meaning, importance, history, and future of trans technologies. *Feminist Media Studies*1-17.
42. Oliver L Haimson, Dykee Gorrell, Denny L Starks, and Zu Weinger. 2020. Designing Trans Technology: Defining Challenges and Envisioning Community-Centered Solutions. In *Proceedings of Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1-13.
43. Oliver L Haimson and Anna Lauren Hoffmann. 2016. Constructing and enforcing" authentic" identity online: Facebook, real names, and non-normative identities. *First Monday*, 21, 6.
44. Aaron Halfaker, R Stuart Geiger, Jonathan T Morgan, and John Riedl. 2013. The rise and decline of an open collaboration system: How Wikipedia's reaction to popularity is causing its decline. *American Behavioral Scientist*, 57, 5: 664-688.
45. Benjamin Hanckel, Son Vivienne, Paul Byron, Brady Robards, and Brendan Churchill. 2019. 'That's not necessarily for them': LGBTIQ+ young people, social media platform affordances and identity curation. *Media, Culture & Society*.
46. Jean Hardy and Silvia Lindtner. 2017. Constructing a desiring user: Discourse, rurality, and design in location-based social networks. In *Proceedings of 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 13-25.
47. Jean Hardy and Stefani Vargas. 2019. Participatory Design and the Future of Rural LGBTQ Communities. In *Proceedings of Companion Publication of the 2019 on Designing Interactive Systems Conference 2019 Companion*, 195-199.
48. Gary W Harper, Pedro A Serrano, Douglas Bruce, and Jose A Bauermeister. 2016. The internet's multiple roles in facilitating the sexual orientation identity development of gay and bisexual male adolescents. *American journal of men's health*, 10, 5: 359-376.
49. Lynne Hillier, Kimberly J Mitchell, and Michele L Ybarra. 2012. The Internet as a safety net: Findings from a series of online focus groups with LGB and non-LGB young people in the United States. *Journal of LGBT Youth*, 9, 3: 225-246.
50. TJ Jourian and Z Nicolazzo. 2017. Bringing our communities to the research table: the liberatory potential of collaborative methodological practices alongside LGBTQ participants. *Educational Action Research*, 25, 4: 594-609.
51. 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 2016 CHI Conference on Human Factors in Computing Systems*, 1152-1156.
52. Sara Kiesler, Robert E. Kraut, Paul Resnick, and Aniket Kittur. 2012. Regulating behavior in online communities. In *Building Successful Online Communities: Evidence-Based Social Design*, (ed.). MIT Press, Cambridge, MA, 125-178.
53. Vanessa Kitzie. 2019. "That looks like me or something i can do": Affordances and constraints in the online identity work of US LGBTQ+ millennials. *Journal of the Association for Information Science and Technology*, 70, 12: 1340-1351.
54. Jes A Koepfler, Katie Shilton, and Kenneth R Fleischmann. 2013. A stake in the issue of homelessness: Identifying values of interest for design in online communities. In *Proceedings of 6th International Conference on Communities and Technologies*, 36-45.
55. Robert E Kraut and Paul Resnick. 2012. *Building successful online communities: Evidence-based social design*. MIT Press, Cambridge, MA.
56. Christopher A Le Dantec, Erika Shehan Poole, and Susan P Wyche. 2009. Values as Lived Experience: Evolving Value Sensitive Design in Support of Value Discovery. In *Proceedings of SIGCHI Conference on Human Factors in Computing Systems*, 1141-1150.
57. Min Kyung Lee, Daniel Kusbit, Anson Kahng, Ji Tae Kim, Xinran Yuan, Allissa Chan, Ritesh Noothigattu, Daniel See, Siheon Lee, and Christos-Alexandros Psomas. 2018. WeBuildAI: Participatory Framework for Fair and Efficient Algorithmic Governance. *Preprint*.

58. Ada Lerner, Helen Yuxun He, Anna Kawakami, Silvia Catherine Zeamer, and Roberto Hoyle. 2020. Privacy and Activism in the Transgender Community. In *Proceedings of Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1-13.
59. Kathryn Macapagal, David A Moskowitz, Dennis H Li, Andrés Carrión, Emily Bettin, Celia B Fisher, and Brian Mustanski. 2018. Hookup app use, sexual behavior, and sexual health among adolescent men who have sex with men in the United States. *Journal of Adolescent Health*, 62, 6: 708-715.
60. Haley MacLeod, Grace Bastin, Leslie S. Liu, Katie Siek, and Kay Connelly. 2017. Be Grateful You Don't Have a Real Disease: Understanding Rare Disease Relationships. In *Proceedings of 2017 CHI Conference on Human Factors in Computing Systems*, 1660-1673.
61. Haley MacLeod, Ben Jelen, Annu Prabhakar, Lora Oehlberg, Katie Siek, and Kay Connelly. 2016. Asynchronous Remote Communities (ARC) for Researching Distributed Populations. In *Proceedings of 10th EAI International Conference on Pervasive Computing Technologies for Healthcare*, 1-8.
62. Juan F. Maestre, Haley MacLeod, Ciabhan L. Connelly, Julia C. Dunbar, Jordan Beck, Katie A. Siek, and Patrick C. Shih. 2018. Defining Through Expansion: Conducting Asynchronous Remote Communities (ARC) Research with Stigmatized Groups. In *Proceedings of 2018 CHI Conference on Human Factors in Computing Systems*, 557.
63. Alice Marwick and Eszter Hargittai. 2018. Nothing to hide, nothing to lose? Incentives and disincentives to sharing information with institutions online. *Information, Communication & Society*, 22, 12: 1697-1713.
64. Adrienne Massanari. 2017. # Gamergate and The Fappening: How Reddit's algorithm, governance, and culture support toxic technocultures. *New Media & Society*, 19, 3: 329-346.
65. Bharat Mehra, Cecelia Merkel, and Ann Peterson Bishop. 2004. The internet for empowerment of minority and marginalized users. *New Media & Society*, 6, 6: 781-802.
66. Alberto Melucci. 1995. The process of collective identity. *Social movements and culture*, 441-63.
67. Cale J Passmore, Max V Birk, and Regan L Mandryk. 2018. The privilege of immersion: Racial and ethnic experiences, perceptions, and beliefs in digital gaming. In *Proceedings of 2018 CHI Conference on Human Factors in Computing Systems*, 383.
68. Guilherme Colucci Pereira and Maria Cecilia Calani Baranauskas. 2018. Codesigning emancipatory systems: a study on mobile applications and lesbian, gay, bisexual, and transgender (LGBT) issues. *SBC Journal on Interactive Systems*, 9, 3: 80-92.
69. Alina Pommeranz, Christian Detweiler, Pascal Wiggers, and Catholijn Jonker. 2012. Elicitation of situated values: need for tools to help stakeholders and designers to reflect and communicate. *Ethics and Information Technology*, 14, 4: 285-303.
70. Annu Sible Prabhakar, Lucia Guerra-Reyes, Vanessa M. Kleinschmidt, Ben Jelen, Haley MacLeod, Kay Connelly, and Katie A. Siek. 2017. Investigating the suitability of the asynchronous, remote, community-based method for pregnant and new mothers. In *Proceedings of Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 4924-4934.
71. Emilee Rader, Kelley Cotter, and Janghee Cho. 2018. Explanations as Mechanisms for Supporting Algorithmic Transparency. In *Proceedings of 2018 CHI Conference on Human Factors in Computing Systems*, 103.
72. Yuqing Ren, Robert Kraut, Sara Kiesler, and Paul Resnick. 2012. Encouraging Commitment in Online Communities. In *Building Successful Online Communities: Evidence-based Social Design*, Robert E Kraut and Paul Resnick (ed.). MIT Press, Cambridge, MA, 77-124.
73. Juana María Rodríguez. 2016. Queer Politics, Bisexual Erasure. *lambda nordica*, 21, 1-2: 169-182.
74. Morgan Klaus Scheuerman, Stacy M Branham, and Foad Hamidi. 2018. Safe spaces and safe places: Unpacking technology-mediated experiences of safety and harm with transgender people. *Proceedings of the ACM on Human-Computer Interaction*, 2, CSCW: 155.
75. Toni Schmader and Constantine Sedikides. 2018. State authenticity as fit to environment: The implications of social identity for fit, authenticity, and self-segregation. *Personality and Social Psychology Review*, 22, 3: 228-259.

76. Aaron Shaw and Benjamin M Hill. 2014. Laboratories of oligarchy? How the iron law extends to peer production. *Journal of Communication*, 64, 2: 215-238.
77. Jesper Simonsen and Morten Hertzum. 2012. Sustained participatory design: Extending the iterative approach. *Design issues*, 28, 3: 10-21.
78. Bongwon Suh, Gregorio Convertino, Ed H Chi, and Peter Pirolli. 2009. The singularity is not near: Slowing growth of Wikipedia. In *Proceedings of 5th International Symposium on Wikis and Open Collaboration*, 8.
79. Joan C Tronto. 1993. *Moral boundaries: A political argument for an ethic of care*. Psychology Press
80. Joan C Tronto. 2013. *Caring democracy: Markets, equality, and justice*. NYU Press
81. Jan E. Trost. 1986. Statistically nonrepresentative stratified sampling: A sampling technique for qualitative studies. *Qualitative sociology*, 9, 1: 54-57.
82. John Vines, Rachel Clarke, Peter Wright, John McCarthy, and Patrick Olivier. 2013. Configuring participation: on how we involve people in design. In *Proceedings of SIGCHI Conference on Human Factors in Computing Systems*, 429-438.
83. Ashley Marie Walker and Michael A DeVito. 2020. "'More gay'fits in better": Intracommunity Power Dynamics and Harms in Online LGBTQ+ Spaces. In *Proceedings of Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1-15.
84. Jillian Weiss. 2011. Reflective paper: GL versus BT: The archaeology of biphobia and transphobia within the US gay and lesbian community. *Journal of Bisexuality*, 11, 4: 498-502.
85. Madisson Whitman, Chien-yi Hsiang, and Kendall Roark. 2018. Potential for participatory big data ethics and algorithm design: a scoping mapping review. In *Proceedings of 15th Participatory Design Conference*, Volume 2, Article 5.
86. Michele L. Ybarra, Kimberly J. Mitchell, Neal A. Palmer, and Sari L. Reisner. 2015. Online social support as a buffer against online and offline peer and sexual victimization among US LGBT and non-LGBT youth. *Child Abuse & Neglect*, 39, 123-136.
87. Kenji Yoshino. 1999. The epistemic contract of bisexual erasure. *Stan. L. Rev.*, 52353.
88. Annuska Zolyomi, Anne Spencer Ross, Arpita Bhattacharya, Lauren Milne, and Sean A Munson. 2018. Values, Identity, and Social Translucence: Neurodiverse Student Teams in Higher Education. In *Proceedings of 2018 CHI Conference on Human Factors in Computing Systems*, 499.

Received June 2020; revised October 2020; accepted January 2021.

Supplemental Materials for Values (Mis)alignment: Exploring Tensions between Platform and LGBTQ+ Community Design Values

Appendix A: Original Conversation Prompts

This appendix provides the full prompt language for each of the design-relevant value elicitation conversations we held with our participants. The general approach here is to use accessible language that is familiar to the population under study (here, the LGBTQ+ community) to frame complex concepts in a useful way.

Conversation 1: Positive/Negative Experiences

To start things off, let's get to know each other.

Please reply to this post and introduce yourself, share your pronouns if you're comfortable, and the meme that you feel like speaks the most to your experiences with LGBTQ+ spaces online.

Then, share your absolute favorite things about being in an online space or using an app made for LGBTQ+ people, as well as your least favorite things about those spaces. This can either be based in specific stories, or it can be a recurring theme you've noticed.

Remember to check back and comment on a couple different or shared experiences, and that everyone's lived experience is going to be unique.

Conversation 2: Blue Sky Wants/Needs

Congratulations - today, you're all officially designers, and the first step in designing something really new is to **think big**, without worrying about costs. For this conversation, we want y'all to tell us about what your ideal online community for LGBTQ+ folks looks like. Here's a few things you could think about to get started, though it's really up to what you think is important:

- How should people be represented or "exist" in the space? Profiles? Avatars? What information should be part of that? And how should those be created?
- How should people connect? Should there be some kind of matching tool? How would you like that to work? Should it happen automatically? Or do you want people to find their own spaces?
- What should the rules or standards be? Should there be rules? If so, who should be in charge of enforcing them?

Once you've written your vision, look at other comments and give feedback on other ideas. **It's important to talk about both the good and the bad when you're designing something, so we'd like you to "yes, and..." at least one other person's vision and "no, but..." at least one person's vision.** Remember, criticism should be both constructive and respectful. If someone posts something that violates our community guidelines, please alert a moderator.

By "yes, and..." we mean finding a vision that you think is great, and helping expand it - maybe to be even better by adding something new, or by showing how your own experiences really support that idea. By "no, but..." we mean finding a vision that you think might not work out the way the person who posted it thinks, and helping them understand why some part of their idea might be a problem for you or folks like you. That "no, but..." could be backed up by your own experiences, and, if you've got a good idea on how to fix the problem, go ahead and suggest it! Feel free to draw on your own experiences and imagination - you're the expert here.

Conversation 3: Algorithmic/Automatic Concerns

Thanks for thinking big for the last two sessions! Now that you've talked about what you want in an online community, it's time to think about how this works. First step: we need to tell the computers what to do. Most online platforms have some kind of automation system that helps make the platform work (sometimes referred to as "algorithms"). For example, we have the system that generates your Facebook news feed, Tinder's matching system, or the automatic sensitive content filters on Tumblr. These systems all handle stuff that may happen too fast and too frequently for humans to handle themselves - but it's up to us to decide how we want these automated systems to work.

First, we have to figure out what criteria the system should use to make a decision; this takes the form of different pieces of data. Data is information that's distinct and measurable. For example, a computer can't make a decision based on someone's "background," because a person's background can be defined many different ways, and is actually made up of many different kinds of information. However, a computer can make decisions on data that roughly add up to that concept - where the person grew up, how old they are, how many years of schooling they had, etc. Then, we have to decide how to rank that data. Ranking tells the computer what data you think is most important to the decision - the higher the rank, the more weight that factor will have in the final decision.

To make this more concrete, we've prepared a few scenarios based on where automatic systems are likely to be used in the visions you worked on the last two days. For each scenario, you need to decide how the automatic system should make its decisions. That means deciding on your criteria/data, and then deciding how you want to prioritize those different pieces of data. We also want to know why you are making the choices you are and want you to explain the tradeoffs you are making. Look through the scenarios and find two you want to try your hand at, then post your criteria/data, ranks/priorities, and a little on why you made these decisions and what you had to wrestle with on the way. Remember to be specific - computers don't handle ambiguity very well at all! Post your design to that scenario's individual thread, not this overall "unit" post.

Once you've posted your two designs, check out how others are designing their systems, and use the yes and/no but technique we used last time to comment on at least two. (They don't have to be the same two you did). Would you change their criteria? Their rankings? Add criteria? Get rid of criteria? We want to know why!

Admitting New Members

There are enough new people who want to join your group that the mod team cannot keep up with vetting each person individually. A system is being implemented that will automatically exclude potential newcomers who seem likely to be trolls or bots. How should the system make the determination about who might be trying to join a group in good faith and who the moderators should not even spend their time vetting? What information should it take into account?

New Content Discovery

Your group is so large and so popular that people can no longer keep up with all of the new content posted to the group, especially content from members they aren't already directly subscribed to. A system is being put in place that shows group members content from across the whole group, not just their existing subscriptions. How should this system pick the new and novel content to show group members? What should it prioritize - content a member has already found interesting? Content that's way outside of what a member might normally see? Content that encourages group members to engage with new people? How should the system decide what fits the goal you pick?

Content Filtering

Collectively, your group has decided that certain types of language or images aren't acceptable in posts for your group. An automatic filter system is being installed to help enforce that decision. How should this automatic filtering system decide what images or words are appropriate and what needs to be filtered - and how should it take the context of the post into account (if at all)? How should the system decide what to remove and what to leave alone? Should it take who is posting it into account - and if so, how?

Tagging

Your online space has enough content generated by the community that it needs to be categorized and sorted. A tagging system allowing people to quickly sort information into relevant categories is being introduced, alongside an automatic tag suggestion system. When you post content, the system will suggest tags - how should it pick the tags to suggest? What criteria should the system use to try and understand and categorize your content? And once the system chooses tags, should they be mandatory, or should you be able to edit them or even opt-out?

Affinity Group/Subcommunity Matching

Your group has gotten large enough and diverse enough that smaller subcommunities based on shared interests have started to emerge. To allow people to find the groups that might be most relevant to their interests and experience, a system that shows group members subcommunities they might be interested in is being introduced. How should this system make decisions about which subcommunities to show which group member? How will the system know when a group member might be interested in a subcommunity? Should the focus be on suggesting subcommunities you would definitely be interested in, or should there be an attempt to introduce some serendipity and variety?

Personal Matching

Though your online community isn't all about dating, hookups, and finding romance, it is something that happens in the group, and a person-to-person matching system with an eye towards romance is being introduced. How should the system decide who you are a "match" with? Should the system assume that opposites attract, or that birds of a feather stay together - or something else entirely? Are there crucial "must haves" or "dealbreakers" the system needs to take into account?

Content Feed

The amount of content posted to your group has increased to the point where most people are having trouble managing the amount of content they've directly subscribed to (e.g., followed). An automatic system is being introduced to determine what content you'll be shown in a "feed" which will be your new homepage. How should this system decide what content to show you, out of the total pool of content you have subscribed to? How should that content get sorted? What's the top priority - what's popular? What's new? Something else entirely?

Conversation 4: Moderation/Policy Concerns

We have been talking about what automated systems can do for our new LGBTQ+ online community space and a lot of the responses point towards wanting more humans to be involved. Today, we want you to tell us more about what you want the humans to do and how. How should moderators be involved in this community? Administrators? What policies are crucial to figure out, how should we figure them out, and how should we enforce them?

Similar to the last conversation, we've put together some likely scenarios, the types of situations which one might encounter in a community like this, based on the type of community you described in conversation two. We want to know how you think the community should deal with each of these scenarios, why you are making the choices you're making, and how you're going to deal with key trade-offs. For example, if we assign duties to humans in the group, we

need to consider how we are going to manage their workload and the possible negative consequences they will face in doing this work.

Same as in the last conversation, you should look through the scenarios and find two you want to try your hand at. Explain how you think the community should deal with the scenario, and explain your thinking, particularly if these scenarios make you question any decisions you made the last conversation. Post your answers to that scenario's individual thread, not this overall "unit" post.

Once you've posted your response to two of the scenarios, check out how others are responding, and use the yes and/no but technique we used last time to comment on at least two. (They don't have to be the same two you did). Would you do things differently? Have different rules? Put responsibility in different people's hands? We want to know why!

Moderator/Admin Selection

Your group has grown large enough that moderators are a necessity. How will you decide who the moderators will be? What are your priorities in picking moderators? Do you try to make the moderator team reflect the opinions, identities, and experiences within the group? How? Should the moderating team have mandatory representation from subgroups that might have specific concerns about norms and dynamics in the group? Also, moderating is a lot of work; how do you make sure that it doesn't always fall on the same people? Are people compensated for their work - and if so, how, and where does the group get the resources for that compensation? Is there an expectation that everyone will take on some minimum commitment to moderating to participate in the group? Finally, how will you, as a group, hold moderators to account, and ensure that the rules are applied as consistently as possible with the least amount of bias? What if a moderator turns out to be a source of problems and bias themselves?

Tagging and Content Warnings

As your group grows, two things are becoming clear: there is too much content for everyone to sort through every day, and some of the topics that people want to talk about are considered unwelcome or problematic by some of your other group members. To deal with these problems, you are implementing a tagging system. What should the focus of this tagging system be - to help people find the content they want to see, to keep people from seeing the content that they do not want to see, or something else? Who should be in charge of doing the work of keeping tagging consistent, and the tags themselves well-organized? Who decides what content should be tagged, and when and if tags should be mandatory? Are content warnings tags, or something special and separate - and are they different in terms of when they are required to be used? Is tagging done as a post is being added to the group or by moderators after the fact?

Education & Onboarding

When new people join your group, it usually takes them a little while to learn the rules and the norms of behavior. How much effort does your group expend in teaching new users about the norms in your group? What does that process look like? Does it include any kind of initial leniency for newcomers - or, conversely, a period of increased scrutiny? Who is in charge of doing this education, keeping in mind that education is a difficult task requiring a lot of emotional labor, and frequently causes burnout?

Community Rules

You are establishing a new group for LGBTQ+ people in your geographic area, and while it is small now, it is reasonable to assume that it will grow larger than just you and your immediate circle. As you are establishing the group, you need to establish what the code of conduct for the group will be - in other words, the rules. Who should be a part of deciding what the rules are going to be? How will you make sure that new people in the group know what the rules are? As your group grows and evolves,, how will you manage rules changes in the future? What are the pros and cons for the set of rules that you choose?

New Member Admissions

Your group is growing, and with that growth comes concerns about trolls and other bad actors trying to join the group to cause trouble. In an effort to keep allowing legitimate members in while screening out people who would be disruptive, the group has collectively decided to have a screening procedure before allowing people to join. Who should be responsible for this screening process, and what criteria should be used to make decisions about who is allowed in the group? Keep in mind that if your criteria are too loose, that will increase the number of people who join the group specifically with the purpose of being destructive. But if you are too restrictive, you will inevitably exclude some members who might genuinely belong in and get real benefits from being part of your group.

Language & Self-Expression

You have a member in your group who is new and is clearly early in the process of coming to terms with and defining their own identity. The group is clearly a rare space where this new person feels comfortable exploring their identity, but in doing so they are using words that are against the language norms in the group. While this person isn't using outright slurs, they are using language that is potentially exclusionary and makes other group members uncomfortable. That said, it is clear that this person is not acting in bad faith, but simply using language they themselves find helpful in the moment. How should the group deal with this situation? Should someone intervene? If so, what should the intervention look like, and who is responsible for it? Does your answer change if this is not the first time this issue has come up with this person?

Extreme Content

Your group is large and well established. There is a ton of new content posted every day. Some of this content includes things like images of violence, sexual assault, and inappropriate images of minors. This content is potentially damaging to the group, traumatizing to members, and puts the group in legal jeopardy. How should the group deal with this kind of content? Who should be responsible for flagging it, and for reviewing/deleting it, keeping in mind that this includes exposure to this kind of content? What kinds of policies and procedures would you put in place for dealing with this content? If humans have to look at and sort through it, how do you account for the risk they're exposing themselves to and the time they are putting in?

Conversation 5: Prioritization & Role-Taking

Over the last week, everyone here has raised important issues and some ways of handling them within online communities. We want to thank all of you for your insight and commitment to talking this out. For our final conversation, let's imagine we're launching this new LGBTQ online community. We're going to launch a platform tomorrow based on the conversations we've had here. What does that platform look like, and **what role do you see yourself playing on this new platform?** What needs do you think this would fill for you - how would it fit into your life?

Remember, your thoughts are valued no matter your level of technical expertise.

Appendix B: Group Rules & Policies

To avoid exposing our participants to unnecessary risks around disclosure, harassment, discussions of sensitive topics, and the harms that frequently impact the LGBTQ+ community, we adopted a team-based moderation approach with a Code of Conduct modeled on codes of conduct from other spaces with similar populations and concerns. General principles and language were drawn from the example policy from the Geek Feminism wiki[1], created by the Ada Initiative and other volunteers, the Working Agreements for Community Cave Chicago[2], and the

[1] https://geekfeminism.wikia.org/wiki/Conference_anti-harassment/Policy
[2] https://communitycavechicago.org/working-agreements

Queer and Asian Conference (QACON) 2019 Community Agreements[3]. Source material was edited and adapted for the research and online contexts, and to ensure the document was concise, easily understandable, and set clear expectations for participants. Here, we provide the code of conduct, our moderation team setup, and our moderation guidelines.

Code of Conduct

Our research group is meant to be a safe and open space for our participants. As such, the group operates with the following code of conduct:

You Know You, I Know Me

Try not to make assumptions about others, related to gender or otherwise. When speaking, please try to use "I" statements and avoid making generalizations or applying your own ideals to others.

What happens here stays here

Though you are welcome to share your own experiences and feelings about the study with others, please refrain from repeating other participants' stories, names, likenesses, etc. outside of the group.

Challenge the idea, not the person

People have a lot of different opinions – and that's great! Disagreement about different priorities is good, and some of what we are trying to learn about here is how different people want to balance those priorities. However, we want to keep discussion centered on those opinions, not the people that have them. If you disagree with an opinion, say so – but don't attack the person.

Oops/Ouch

If something offensive, problematic, or hurtful is said or done in the group, anyone may say, "ouch." The person that had been speaking should please say, "oops," and then the problems with what was said should be discussed by those persons and/or the group.

Ouch, Anon

If any person feels that an "ouch" needs to be said, but is not comfortable saying so at the moment of occurrence, this should be communicated to our moderators. If you are comfortable identifying yourself, DM one of the study team members. If you wish to report anonymously, use the reporting form, which will send an anonymous report to our moderator channel.

Don't Giggle My Wiggle

Folks here have different tastes and preferences, so avoid antagonizing language like "I hate that," or "ew." Likewise, folks have different traumas and triggers, so avoid language that belittles or trivializes their experiences.

Harassment

We are dedicated to providing a harassment-free experience for everyone. We do not tolerate harassment of participants in any form. Participants violating these rules may be removed from the study at the discretion of study staff. Refer to the moderation guidelines for more information. Harassment includes, but is not limited to:

- Comments that target other participants based on characteristics such as gender, gender identity and expression, sexual orientation, race, ethnicity, age, ability status, physical appearance, body size, or religion.
- Deliberate intimidation, stalking, or following
- Unwelcome personal attention

[3] https://qacon.org/qaconchecklist.html

- Persistent, unwanted attempts to contact another study member
- Advocating for, or encouraging, any of the above behavior

Moderation Guidelines

If a violation of our code of conduct occurs, we follow a three-level procedure for dealing with incidents:

Level 1

Participants are encouraged to first respond to posts or responses they find problematic by employing the "Oops/Ouch" principle from our working agreements. This is especially true in cases where the intent is clearly not expressly to offend. If you are comfortable, participants are encouraged to post a short response to the comment in question indicating that you would prefer folks to avoid that type of posting and why, then lead the topic gently back in the right direction with some substantive comment on the subject matter in discussion. In cases where offense appears to be the intent, participants are encouraged to escalate to the "Ouch, Anon" principle.

Level 2

In the case of a report from a participant (as laid out in the "Ouch, Anon" principle), or a case of obvious malicious trolling or hate speech, moderators will review the post in question and, if appropriate, record the content of the post for future analysis and remove the original from the thread. The moderator will notify the participant of this privately via direct message and explain how the response is not within the group guidelines, requesting that further responses of that nature not be entered in to the group conversation.

Level 3

In the case of repeated violation of our working agreements (e.g., 3+ incidents), a project co-investigator/administrator will make a decision as to the offending participants continued participation in the research community. This decision will largely be based on the participant's effect on the ongoing safety and norms of openness for the group as a whole. Repeated offenders may be asked to leave the group as a last resort, and only after following the steps outlined in the procedures above have been followed. By the time a participant is banned, it should have been made very clear to them that they are behaving unacceptably and have been informed of the terms of continued participation before they are banned. Being asked to leave the group will not require the offending participant to forfeit their initial payment for participating in the study, however they will not be allowed the opportunity to participate in the follow-up interviews and subsequent payment.

Moderation Team Setup

We actively moderated the research group from 8 am to 8 pm, US Central Time, and allowed participants to reach out via our anonymous reporting form or Facebook direct message at any time. During the duty hours, several roles were always assigned as on-duty, which in some cases doubled as analysis roles to support making future conversations more responsive to the results of past conversations.

Active Moderator

While on duty, active moderator should spend most of their time in the group itself. The first priority is to ensure that the code of conduct is being followed within the group, and all participants are experiencing a productive, "brave" atmosphere. This does not mean banning and suspending people as primary tools, and anyone in this role should be following the moderation escalation procedures spelled out in the code of conduct. Rather, this role is about stepping into conversations before they turn irreparably nasty, and turning them back towards subjects that

support answering our research questions. De-escalation is the key here, especially through reminding people to focus on/debate on ideas instead of ad hominem attacks. When not carrying out these safety/security tasks, the active moderator's job shifts to prompting discussion and asking follow-up questions. As a first pass, look for vague responses that could use expansion in order to help us understand what's being said and make it easier for other people to respond to. As a second pass, ask substantive follow-up questions that push on assumptions. *Note: first shift active mod is responsible for doing all of this for whatever came in overnight.*

Analysis Lead

The analysis lead's primary job is to be pulling out themes and points of interest from the data as it develops, comparing this back to previous data in order to build up larger themes that cross conversations. While on duty for this position, focus on using your Research Log as a memo in which you lay out what you are seeing and try to make connections to our larger research questions. Pay special attention to any information that will help us set up the next conversation (e.g., in conversation one, start trying to pull out the values being revealed so we can sum them up to set up conversation two). When useful for further analysis, the analysis lead should also go into the group itself and ask follow-up questions in the appropriate threads. As a secondary duty, the analysis lead backs up the active moderator, helping deal with any issues within the group in real time. *Note: first shift analysis lead is responsible for helping the first shift active moderator clear any backlog and start on analysis from overnight responses.*

On Call

The two call spots are essentially backup for times when the group has heavy traffic, or if a serious incident occurs. The call spots should check in with the active moderator periodically to see if they can help. Beyond this, they should focus on whatever secondary tasks have been assigned (e.g., working on engagement materials or small tasks for other ongoing projects) or, if no specific assignment has been made, continue general analysis via their Research Log.