How Transfeminine TikTok Creators Navigate the Algorithmic Trap of Visibility Via Folk Theorization

Michael Ann DeVito, University of Colorado Boulder

Social platforms open important doors to visibility for transgender people, through which they can pursue key goals such as broader recognition and normalization. However, each door is also potentially a trap, filled with risks and consequences — especially for those whose goals require visibility. Via a grounded theory interview study with 17 transfeminine content creators on TikTok, I find that, in an algorithmically mediated environment such as TikTok, users navigate potential doors to visibility and their associated traps via folk theorization. Moreover, I find that transfeminine creators employ multiple complex and overlapping folk theories, with *actionable theories* guiding the careful navigation of doors to visibility, and *demotivational theories* alerting creators to traps that are too risky to spring. I introduce five novel folk theories of TikTok spanning both the For You Page and content moderation systems which creators use to guide their decision making, and discuss how two cross-cutting issues, perceived *algorithmic paternalism* and decontextualization, illustrate major issues for transfeminine creators and opportunities for more supportive design.

CCS Concepts: • **Social and professional topics → User characteristics → Gender;** Human-centered computing → Human computer interaction (HCI) → Empirical studies in HCI

KEYWORDS

Transgender, transfeminine, folk theories, TikTok, content creation, content moderation, harassment

ACM Reference format:

Michael Ann DeVito. 2022. How Transfeminine TikTok Creators Navigate the Algorithmic Trap of Visibility Via Folk Theorization. In *Proceedings of the ACM on Human-Computer Interaction*, in press.

1 INTRODUCTION

She knew she was going to be harassed if she kept posting. More and more hateful comments were piling up on her videos as she gained followers and reached new audiences; more and more transphobic ones, too. She tried turning on comment filters, but that was never effective—it seemed like the system missed the transphobes and marked the people who needed help as hateful. She considered not posting anymore—after all, it wasn't her job. Yet still, she thought, how much would it have improved her life to see someone like her when she was young, and to ask them questions? To see a transgender woman, not just surviving and embracing societal stereotypes of femininity, but thriving and succeeding in her own distinctly-nerdy niche of femme identity? To see that, if she came out as transfeminine[2], there was still a good future for her? It would have meant the world.

---

[2] In this paper, I will use the term "transfeminine" to refer to individuals who were assigned a male gender identity at birth, but currently identify with or hold gender identities that our larger society considers feminine. This group specifically includes transgender women and transfeminine nonbinary individuals, all of whom are primarily affected by transmisogyny [38, 46, 47]. Throughout the paper, the terms "transfemmes" and "transfeminine individuals/creators" can be read as synonyms.

In the vignette above, a transfeminine content creator on TikTok faces a common yet challenging strategic decision: pursue visibility to achieve her goal of helping other transfemmes, or play it safe and avoid the risk of harassment and abuse. Gossett et al. and Pinter et al. have theorized that while visibility is, on many fronts, a primary path forward for trans people, a *door* to progress, it is also always a potentially-harmful *trap* in that the acceptable bounds of visibility are limited, and any visibility brings new risks [28, 41].

For transfeminine people, all visibility is likely to be a trap. Transfemmes are uniquely impacted by *transmisogyny*, the intersection of transphobia and misogyny that leads to the devaluing, oversexualization, and cultural hatred of trans women and transfeminine nonbinary people [46]. As such, transfeminine people pursuing visibility online are likely to be targeted by the increased misogynistic harassment and inequitable rule enforcement experienced by women [5, 18, 20, 39], the targeted, transphobic harassment and exclusion experienced by transgender people broadly [44], and the additional, unique oppressions that come from existing at the intersection of these two categories [38]. Within an environment where harassment is almost a given, transfeminine creators must make strategic decisions about which doors to visibility they open, assessing the risk of springing the related traps against the fact that visibility is necessary to pursue their goals. Moreover, in the context of TikTok (and most algorithmically-based social platforms), transfeminine creators must make these decisions in an opaque algorithmic environment [14, 15, 26, 33].

Past work on user understandings of opaque, algorithmically-driven social platforms suggests that the likely mechanism behind this decision making is folk theorization, or the formation of informal, socially-constructed, affect-influenced conceptions of how a platform works [13, 14, 15, 22, 23]. Indeed, there is already evidence that folk theorization is in play on TikTok more broadly, and impacts how users attempt to shape their primary content feed, the For You Page (FYP) [33]. Past work has also established that TikTok users theorize with identity in mind [33, 48]— the very thing that transmisogyny seeks to invalidate for transfemmes [46].

In this paper, I will build on Gossett et al. and Pinter et al. to argue that visibility on TikTok, while crucial for the pro-social, educational, and community-building goals of transfemme creators, is also always a trap due to design and policy decisions which heighten the risk of harassment, transmisogyny, and perceived suppression. Based on a grounded theory interview study with 17 transfeminine TikTok creators, all of whom actively experience harm due to their visibility, I extend the trap of visibility lens to the space of algorithmically-driven social platforms and visibility mechanisms. Using this lens, I will argue that transfeminine content creators rely on multiple, overlapping folk theories of the platform and its components to assess the risks and benefits of the platform's multiple offers of visibility in order to keep themselves safe while still accomplishing their goals. I also extend the current understanding of folk theorization in social computing to include **actionable folk theories**, which enable creators to adapt platform issues and still achieve their goals, and **demotivational folk theories,** which leave creators with no clear path forward and erode positive perceptions of platform spirit. Finally, I identify five novel folk theories which participants employed in navigating algorithmic traps of visibility: the **algorithmic community** theory, the **expanding stages** theory, the **algorithmic paternalism** theory, the **identity flattening** theory, and the **cultural context** theory.

2 BACKGROUND

Visibility is a complex subject for transgender people and, as Gossett et al. argue, always both a door to potential progress and a trap waiting to be sprung. Gossett et al. detail how increased visibility of transgender identity is held out as "the primary path by which trans people might have access to livable lives," as it potentially paves the way for increased recognition, understanding, and acceptance from cisgender society, and the resources this understanding

brings [28]. Additionally, there is also evidence that being open about one's identity has long-term health benefits for the individual [53]. However, visibility is also potentially harmful, as increased visibility can lead to increased stigmatization, harassment, and harm, even from within the LGBTQ+ community [44, 52]. Moreover, the positives of visibility are not equally distributed, as they are only applicable to those that are within a publicly "acceptable" range of transgender identity in terms of appearance, behavior, economic status, and race [28].

Gossett et al. propose that the complex relationship between visibility and transgender identity can be captured via the lens of the *trap of visibility*, which focuses on *doors* to visibility, the unexpected and harmful *traps* which lie behind each door, and the *trapdoors* trans people employ to find alternative forms of visibility. While Gossett et al. focus on the trap of visibility as a way to analyze visual media and culture at a societal level, Pinter et al. have extended this lens into social computing as a way to analyze the individual coming out experiences of transgender people in the context of groups on algorithmically driven social platforms. Importantly, they argue that, when making decisions about how to behave online, trans people carefully evaluate both doors and their associated traps [41]. Here, I build on both versions of the lens in order to apply it to the public and algorithmically-mediated context of content creation on platforms like TikTok.

Doors, in the Gossett et al. sense, are moments when trans people are offered specific opportunities for visibility and the potential positives which come along with it, having previously been excluded from visibility [28]. In the context of creators making decisions on social platforms, these offers of visibility are instantiated into affordances and feature sets. For Pinter et al., doors are "the exterior representation of a community" in which an individual might decide to make their trans identity visible. In the current context, these are perceptions of what a platform has on offer, likely gleaned through a combination of observing platform behavior and discussing the platform with others [13]. As such, I define **algorithmic doors** as *algorithmically-driven platform affordances and features which are perceived to offer increased content and individual visibility of marginalized identity if used.*

Traps, in the Gossett et al. sense, are the societally-informed, externally-imposed bounds on acceptable trans visibility which exist behind and parameterize each door, acting to ensure that the door to visibility is only used for the presentation of trans identities that those with societal power find acceptable and nonthreatening [28]. Traps remind us that doors to visibility are not magic portals to any kind of visibility — like walking through a physical door, accepting a door to visibility brings the user to what is on the other side of *that specific* door, not *any* door. Doors to trans visibility lead to opportunities for visibility within a strictly-bounded, identity-specifying box, and using that visibility to communicate identity outside of those bounds often results in increased stigmatization and lack of support from those offering the door to visibility [28]. In the current context, these "acceptable bounds" for trans identity are still societally informed but are instantiated through platform policy and mechanisms. The platform's content distribution systems set criteria for which pieces of content get what type of distribution, and the platform's content moderation systems and body of policy decisions set criteria for what types of visibility (and identities) are eligible for continued protection under hate speech and harassment policies. Platforms define an acceptable algorithmic identity for the trans user and assign it in place of the user's actual identity, as Cheney-Lippold describes [9]. Then, the platform effectively requires trans users to infer what that algorithmically-imposed identity is and what boundaries it imposes, and stay within those bounds or face consequences. As such, I define **algorithmic traps** as *system design decisions, policy decisions, and practical enforcement realities that bound acceptable uses of and reactions to uses of algorithmic doors to visibility along identity and behavioral lines, limiting the ongoing safety and effectiveness of said doors.* Algorithmic doors are not offers of visibility for any kind of trans identity — they are offers of visibility for specific trans identities which may or may not actually align with a user's core identity or the identity it is their goal to present, and the algorithmic traps associated with each

door instantiate this bounded standard of acceptable identity into algorithmically-driven content delivery and moderation decisions at scale. Notably, following Pinter et al., this work views traps as fluid over time [41], as platforms constantly change in ways that can make it difficult for users to predict or adapt to their behavior [13, 25].

Trapdoors, in the Gossett et al. sense, are "clever" alternatives to accepting the traps behind each door, passageways which lead to new, usually unknown options [28]. Gossett et al. conceptualize these trapdoors broadly, and actively call for the invention of new, novel, and unexpected trapdoors. Pinter et al. operationalize this concept as the workarounds (and possible future workarounds) which users employ to avoid or escape from traps [41]. In the current context, I define **algorithmic trapdoors** as *strategies and tactics which individuals or communities use to avoid, modify, or counter the unwanted outcomes of allowing one's content and identity to be mediated by an algorithmically-driven social platform's systems.*

Ultimately, though they thoroughly recognize visibility as a trap first and foremost, Gossett et al. argue that doors to visibility can still be very useful — so long as the trans person walking through that door understands the nature of the trap they are springing, or can find a useful *trapdoor* by which they can essentially escape the trap [28]. Pinter et al. found evidence that there was ample reason for trans users to want to go through potentially-risky doors to visibility, including increased social support and the ability to help other trans folk in turn [41]. Moreover, they found that by going through these doors, trans users were able to essentially reinforce and hold these doors open for others, making it easier for other trans people to see the door, evaluate the traps behind it, and go through it if they so choose [41].

Pinter et al. call for work using the trap of visibility lens in different contexts and with participants who have experienced direct harms, instead of participants who feared potential harms (the focus of their study) [41]. Here, I answer that call by focusing on transfeminine creators, a group in which I found direct harms to be common, deliberately pursuing visibility on TikTok. In the remainder of this section, I will detail how this new context varies in important ways from past work, how theorization and sensemaking are likely happening in this context, and what we already know about TikTok and stigmatized identities.

2.1 Deliberate Visibility, Immediate Risks: Transfeminine Experiences as Context

Queer and trans identities are not monolithic in their composition or the oppressions they face [44, 52]. The experiences of and oppressions experienced by transfeminine people are at least partially unique to transfemmes [38, 46, 47], and are tied to misogynistic oppressions against women [38, 46]. Transfeminine people are uniquely affected by *transmisogyny*, which Serano defines as the unique form of oppression that results from being subject to both misogyny and transphobia [46]. As such, transfemmes can expect to be subject to harassment, abuse, and stigmatization from at least three sources: 1) the sexism-based problems faced by women (e.g., uneven enforcement of platform rules [20], extensive harassment [5, 18] with lasting consequences when the harassment is identity-based [39]); 2) the transphobia-based problems faced by trans people online (e.g., abuse and stigmatization [44], harms due to exclusion from design consideration [2]); and 3) the unique oppressions which emerge at the intersection of the two [38, 46]. Indeed, past work shows us key differences in minority stress levels and ability to be resilient among different types of trans people [42]. Even in comparison to other transgender people, transfemmes face more severe economic consequences when they are open about their identities [45]. Considering that algorithmic traps bound acceptable visibility along identity lines, the challenges faced by transfemmes require specific consideration.

It is also important to note that, in social computing, the tradeoffs between the benefits of trans visibility and the dangers of harassment and stigmatization have mostly been explored in the content of individual visibility, coming out, and gender transition. General LGBTQ+ work

exploring this topic has largely focused on ensuring selective visibility of an individual's LGBTQ+ identity [6] and preventing context collapse that could lead to stigmatization in a personal context [6, 16, 19], weighing these risks against social benefits [24]. Notably, trans-specific work generally focuses on this personal context as well, documenting the tradeoffs between the identity development, social support, and health benefits of being visible as trans on social platforms during key life stages such as coming out and transitioning, and the potential pitfalls of being visible to the wrong people, or at the wrong times, while reaping these benefits [30, 31, 41, 44]. Much of this work also focuses on selective disclosure for the individual, especially in contexts where the participant is not fully, publicly out as transgender.

In this study, my focus is not on the personal context of selective disclosure, but rather on the context of publicly managing an explicitly-transfemme identity. As I will demonstrate in section 4.1, the transfemme creators in this study are publicly out and have no desire to conceal that they are trans, and their mission is often explicitly educational, not primarily personal. This leaves the common solution to the possibility of transphobic harassment and abuse — avoiding or extensively segmenting disclosures of trans identity [6, 16, 41, 44, 52] — off the table. Selective disclosure of trans identity is not the goal for transfemme content creators; rather, specific audience targeting is. This puts transfemme TikTok content creators in a space closer to more broadcast-style platforms than traditional social media — a space where harassment and abuse are the price for becoming visible, particularly for LGBTQ+ individuals and women [51].

2.2 Adaptation and Folk Theories

While the harassment, stigmatization, and abuse situation appears grim for transfeminine creators, that is not to say that people cannot or do not adapt to and around these problems. In the personal contexts discussed above, it is common for trans users to use affordance- and ecosystem-based strategies to ensure selective visibility, as well as non-technical solutions such as modifying the extent of trans visibility in their content and steganography [6, 16, 41, 44, 52]. While there is little work on trans social platform users outside this personal context, we can see that women and LGBTQ+ people more broadly also adapt, often via privacy affordances and the extensive use of moderation tools [51].

In the broader context of algorithmically-driven social platforms and the pursuit of content visibility, a growing body of work argues that adaptation to and around the algorithmic components of platforms is driven by folk theorization [13, 14, 15, 22, 23, 43]. Folk theories are defined by DeVito et al. as "intuitive, informal theories that individuals develop to explain the outcomes, effects, or consequences of technological systems, which guide reactions to and behavior towards said systems" [15 p. 3165]. Subsequent work has updated this definition to characterize folk theories as socially-influenced and malleable [14]. It has also moved the focus of study from folk theories of a single algorithm (e.g., the TikTok FYP) to algorithmic platforms as a whole (e.g., the FYP as well as the content moderation, advertising, and reporting systems), recognizing that folk theories have both mechanistic and affective components [13] when talking about adaptation specifically.

Folk theorization of algorithmic platforms is a constant loop in which users forage for information both via observation and testing on the platform itself and social sources, and use this information for sensemaking, which updates one's folk theory, which in turn guides one's behavior, generating new information to take into account in future folk theorizing [14]. This process takes place within and is impacted by both the individual goals of the theorizer [14] and by the theorizer's sense of the platform's *spirit* — "the user's perception of what a platform is and what it is for, as determined by the user's understanding of the platform's stated mission, its values and actions in practice over time, and the functionality which it allows as juxtaposed with the user's understanding of the platform's purpose" [13]. The user's perception of the platform's spirit is a major factor in determining if the user will put in the effort to adapt to an

algorithmic system, with positive perceptions of platform spirit generally prompting the user to continue to try to adapt, and negative perceptions of platform spirit motivating decisions to reduce or stop use of a system [13].

The complexity and utility of these theories varies. Early work on folk theorization suggested that baseline algorithmic awareness was low, and that user folk theories were mostly abstract, focused on what a system does and why, not how [22, 23, 43] — what I have previously referred to as *functional folk theorization* [13]. Subsequent work has found increasing amounts of algorithmic awareness and complex folk theorization that addresses the mechanisms behind the platform [11, 14, 15, 36] — what I refer to as as *structural folk theorization* [13]. I have advocated for finding ways to help users move from functional to structural folk theorization, arguing that the knowledge of mechanism sets users up to productively experiment and adapt while maintaining their own autonomy and achieving their goals — though I have also specifically noted that there is no such thing as a "bad" or "good" folk theory, only theories that help one achieve their goals and theories that do not [13].

2.3 TikTok & Algorithmic Representational Harm

Regardless of the overall prevalence of folk theorization, recent work on TikTok specifically suggests that not just folk theorization, but relatively complex, structural-level folk theorization is common among users of the platform [33]. Multiple studies of what Klug et al. call "produsers," or average users of the platform who consume content and make some as well, have found that there is an unusually high amount of algorithmic awareness on TikTok [33, 37, 48]. Karizat et al. found that TikTok users are not only aware of the algorithmic nature of the platform, but are actively folk theorizing about content delivery via the FYP [33].

Both Karizat et al. and Simpson & Semaan found that theorization in this context was heavily tied to identity, with Simpson & Semaan finding that users felt the platform was sorting people into "algorithmic identity categories" [48]. Karizat et al. encapsulated this in what they call the *identity strainer folk theory*, with "users believing that their social feeds are the result of an algorithm recognizing, classifying, sorting, and suppressing social identities based on its conception of which social identities are (or are not) 'valuable' and 'wanted,' or which ones (do not) deserve visibility" [33]. They assert that this is a form of *algorithmic representational harm*, or harms users face from being effectively erased by the algorithm due to a lack of algorithmic privilege [33]. Moreover, Simpson and Semaan specifically point to LGBTQ+ users as being affected by this kind of erasure and exclusion, even among different identities within the LGBTQ+ umbrella [48].

Considering the likelihood of folk theorization in this context, juxtaposed with the trap of visibility that transfeminine creators face, in this paper I ask:

- **How do transfeminine content creators on TikTok with visibility-requiring goals navigate the trap of visibility?**
- **What folk theories and folk theorization do transfeminine content creators have of TikTok, and how does this impact their strategic decision making around visibility?**

3 METHODS

To investigate these questions, I conducted an interview study with 17 transfeminine TikTok creators, using a constructivist grounded theory approach [7, 8]. I approached this work as a complete-member-researcher who is fully embedded in the context under study and immersed in the same phenomena as the participants [1], allowing increased trust, openness, and access during the research process while also providing a deep contextual basis for analysis [1, 21]. This study was approved by my institution's Institutional Review Board.

3.1 Participants

As past work has documented, many varied users of TikTok face situations where they feel their identity is underrepresented on the platform [33, 48]. Here, I choose to focus on transfeminine creators specifically. When doing folk theory work and eliciting latent knowledge, it is often useful to focus on a group with heightened concerns for both ease of elicitation and the ability to capture both mainstream and particularly challenging outlier experiences [13]. Transfeminine people are an excellent example of one of these heightened groups in the context of theorization about social platforms, and especially platform moderation systems, due to the factors discussed in section 2.1. As such, transfemmes are in an ideal position to, through sharing their informal theorization and lived experiences on a platform, help generate close-to-the-ground desubjugated knowledge [50] on this topic.

In a similar vein, I chose to focus exclusively on people who primarily identify as content creators. Past work has covered mixed-use groups on TikTok (e.g., [33, 37, 48]), and by focusing on creators, I can more thoroughly investigate the problems faced during content creation specifically, with a group of participants who are more likely to be aware of these issues due to their own use goals.

3.1.1 Recruiting

Recruiting took place via TikTok itself. I posted a recruiting video[3] with study details and a request to share and engage with the video to my own TikTok page in early February 2021, after two months of posting trans and transfemme-specific content to my account and engaging with other transfemme creators. At the time, I had 3,208 followers. The video was shared 729 times and was viewed by over 20,100 TikTok users, 72% of whom saw it on the FYP, suggesting they were from outside my immediate network. To broaden distribution, I added the hashtags #transtok, #algorithm, #algorithmtiktok, #trans, #transfemme, #girlslikeus, #lgbt, #mtf, #transgirl, and #research to the post. Interested viewers were directed to direct message me, so long as they fit the eligibility criteria: transfeminine, 18 or older, a content creator, and over 1,000 followers to ensure that the participant was actively creating content and pursuing a following[4].

3.1.2 Demographics

To preserve participant privacy, I do not report on individual identities or demographics. My participants ranged in age from 19 to 52 years old (M=29). All participants were content creators posting to TikTok, with followings ranging from 1,431 to 40,500 followers at the time of initial interview (median=8656 followers) and content that had drawn from 10,600 to 1,100,000 likes at the time of initial interview (median=157,200 likes). A range of genders under the transfeminine umbrella are represented, including creators who are binary trans women, transfeminine nonbinary, nonbinary trans women, and transfeminine. Similarly, a range of sexual and/or romantic orientations is represented, including lesbian, bisexual, pansexual, queer, "enby for enby,"[5] and androsexual[6]. All participants were out about their gender and sexuality, especially in an online context. Participants variously use she/her, she/they, and they/them pronouns; participants who use multiple pronouns will be referred to using both pronouns in the rest of this paper.

---

[3] https://vm.tiktok.com/TTPd9mYxLb/

[4] 1,000 followers is a key milestone on TikTok, as it unlocks the ability to go live. Additionally, the primary creator organization for TikTok, The Online Creators Association, has established this as the threshold for being an official creator and admission to the association.

[5] An orientation in which a nonbinary person is primarily attracted to other nonbinary people.

[6] An orientation in which a person is primarily attracted to those who present in a masculine way.

In terms of racial identity, participants were majority white; however, the sample also includes Arab, Desi, and Black transfemme creators. Additionally, this sample includes neurodivergent individuals (both ADHD and Autistic neurodivergence) and Jewish individuals who must respectively contend with ableism and antisemitism in addition to transmisogyny. Participant occupations, outside of content creation, include undergraduate student, graduate student, service worker, educator, medical professional, graphic designer, technical support personnel, and software engineer. Participants represent a variety of educational and economic backgrounds, ranging from high school diplomas to advanced degrees and self-described "economically disadvantaged" to middle class, respectively. Only two participants had educational backgrounds that provided formal knowledge of algorithms, and their responses were, while more detailed in their speculation, in line with participants without this training. All participants were located either in the United States or the United Kingdom.

3.2 Interviews

Initial interviews took place throughout February 2021 and were conducted via Zoom, video and audio recorded, and auto-transcribed by Zoom's transcription service. The interview period started with the informed consent process. Initial interviews ranged from 70 to 115 minutes (M=87 minutes, SD=13 minutes).

I took a semi-structured approach to questioning, covering topic areas including: identity, self-presentation, and history; content creation goals, style, tactics, and problems; attitudes and history regarding TikTok; and folk theorization. The folk theorization component of the interviews included a whiteboard drawing exercise via Google Jamboard in which participants were asked to make an image of how TikTok distributes their content, as visual exercises are a key method for eliciting folk theorization information [13, 14]. My interview topic guide and detailed instructions for the drawing exercise are included in this paper's supplemental materials. I used the image as a starting point for the folk theorization portion of the interview, stepping through the image with the participant and probing on folk theories, sources of folk theory information, change over time, adaptation, and the impacts of adaptation. Throughout the interview period, I adjusted and updated my questions as part of my initial theoretical sampling [7, 8].

After preliminary analysis, I conducted follow-up interviews with nine participants in order to further my theoretical sampling and develop emerging theoretical categories [7, 8]. Follow-up interviews took place in mid-March 2021 and ranged from 27 to 71 minutes (M=51 minutes, SD=17 minutes).

3.3 Analysis

I took a constructivist grounded theory approach in this study [7, 8], engaging in open coding and memoing throughout data collection to assist in theoretical sampling. This was followed by an analysis period that included additional memoing and open coding, followed by a focused coding round, and then a theoretical coding round in which I related codes to each other and to applicable existing theory. I used MAXQDA as my analysis platform.

3.3.1 Positionality

Positionality necessarily informs analysis; as such, my position as sole investigator and author is both a key benefit and an important limitation to this study. I am a complete-member-researcher for this population [1], in that I am transfeminine, and an active content creator on TikTok who, as of the time of writing, has 10,216 followers and 231,012 likes on my content. I have personally experienced many of the issues described by participants, including extensive transphobic and transmisogynistic harassment on TikTok, ranging from deliberate misgendering to repeated, graphically-detailed death threats. Via lived experience and

actionable folk theories help creators evaluate doors and navigate traps in service of adaptation, while **demotivational folk theories** steer creators away from platform tools while eroding positive perceptions of platform spirit.

4.1 Promising Doors: Positive Perceptions of Platform Spirit

Algorithmic doors to visibility can be evaluated at both the high level of the whole platform and the low level of specific mechanisms and options within the platform. The existence of the TikTok platform itself can be viewed as one big offer of visibility and evaluated as such, while the practical options for pursuing visibility on the platform can be viewed as individual offers of visibility which can be separately evaluated. In this section, I will discuss how creators approach and evaluate the door to visibility offered by the entire platform; doors to specific paths to visibility will be covered in sections 4.2 and 4.3.

In the context of online platforms, the transfeminine creators in this study appear to employ their perception of a platform's spirit as a tool to assess TikTok's high-level, overall offer of visibility and the associated traps. As I have noted in prior work, spirit has multiple components, the most important in this case being overall history with the platform and task/technology fit [13]. My findings suggest that TikTok is initially perceived as having positive platform spirit due to unusually on-point task/technology fit and the creation of algorithmic community, encouraging creators to step through the door to the entire platform.

4.1.1 Unusually Snug Task/Technology Fit

Across all my interviews, participants consistently discussed how well the basic creator tools and overall environment of TikTok allowed them to pursue their goals as creators, suggesting that TikTok has excellent task/technology fit. The task, in a folk theory context, is mapped to the overall goals participants had when posting to (and theorizing about) TikTok [13]. Here, that goal was largely to help. While some participants did mention seeking validation of their identity via the platform, all participants had some sort of primary educational or social mission which motivated their posting. Most participants wanted to help other trans people, especially those still discovering their identity, and knew that getting onto the FYP would be a key component of achieving that goal, as P2 notes:

> I want my videos to land on the FYP of a trans person who's struggling with coming out. I want my videos to land on the FYP if someone who is struggling with accepting their own reality and accepting who they are and loving themselves. I want my videos to help people.

Participants saw visibility on TikTok as a way to augment or correct what they saw as incorrect or limited depictions of trans people, and especially transfemmes, in existing media and online spaces. For P3, that means simply showing that a trans existence can be a happy one:

> I didn't think that being trans and being happy could go hand in hand before coming out and before starting this transition journey, because I had never seen a happy trans woman. I had seen TV where trans women are miserable or murdered usually, and that sucks... I just was so miserable living as a boy, that I was like, "fuck it, I'll take all of the negative to not have to do this anymore," and the fact that there was positives kind of blew me away. I wanted other people to be able to see that.

Other participants felt similarly, targeting their efforts towards showing the possible diversity of transfemme experiences in a media environment that only shows a limited set of trans identities. For example, P11 had not encountered portrayals of trans people of faith and wanted to provide that visibility for other trans people. P4 noted the importance of showing that trans women could continue to enjoy and participate in sports. P17 drew contrasts to the

4.2.1 Hashtags as Doors with Navigable Traps

For creators operating with an expanding stages folk theory, one particularly important algorithmic door to visibility to evaluate is the hashtags which accompany the initial posting of a video. Choosing hashtags is non-trivial for most participants, as they recognize it as important in terms of getting through the first "gate" to wider distribution, as P4 notes:

> I think that hashtags are useful to begin to kind of escape that first level of content... because there does seem to be this thing where there's a threshold to escape the bounds of a video that doesn't do well, and I find a lot of times that hashtags seem to be pretty helpful in getting there.

P4, like most participants, realizes that hashtags are particularly crucial to gaining visibility outside one's existing following. As they are heavily involved in sports, they often use topical hashtags (e.g., #skiing, #cycling) to increase the chance of distribution to others who are interested in those sports. However, P4 sees and has experienced the algorithmic trap behind this door: wider topical hashtagging appears to be processed by TikTok with the embedded assumption that people who are interested in a topic will be interested in content on that topic from creators of any identity, and will behave appropriately when they encounter it. Practically, this results in a noticeable uptick in transmisogynistic engagement for those that employ topical hashtags. For some, though, this is a trap worth risking, as P9 explains:

> I don't want to keep from tagging [content topics] because I want to bring in people from those interests, and if other people get brought in too, then I'll deal with that, but I don't want to let transphobes ruin anything for me.

For P9, who is explicitly trying to make a broader range of transfemme identities visible, it is worth it to risk transmisogynistic engagement, as broadening her reach via topical hashtagging increases her visibility to those with different interests that may not generally be assumed to be common for transfeminine people. Moreover, it heightens the likelihood that transfeminine people who have not yet been exposed to specifically-trans content or the trans "side" of TikTok will see the video. She is aware of the consequences of this algorithmic trap — more visibility will bring more abuse — but proceeds knowingly, prepared for the consequences.

Of course, not all participants used hashtags in this way. In fact, some see the algorithmic trap attached to topical hashtags as too risky, and use protective, identity-focused hashtags to help them achieve only the visibility they want, while avoiding the risks of wide distribution, as P5 explains:

> I never let it go broad cuz I don't want to end up on straight TikTok. From the start, I would tag it with literally every single queer hashtag like queer, bisexual, pansexual, lesbian, gay, trans, asexual. I would hit them all just so that it would stay within the community I wanted to hit. Some videos that I forgot to tag with all of those did end up on straight TikTok and those videos got stressful.

P5 sticks to queer hashtags exclusively, as she is primarily interested in showing other trans people that transfemmes can have good, fulfilling lives — for her, directly employing topical hashtagging is too risky considering her goals. Others, such as P3, combined approaches to create an algorithmic trapdoor, essentially using identity-focused hashtags additively with topical hashtags to attempt to steer content distribution. P3 enjoys posting videos about cryptids, but notes that the cryptid hashtags are often full of conspiracy theories and transphobia, and use of the #cryptid hashtag alone generally attracted transmisogynistic engagement. As such, she now also uses a slew of queer-specific hashtags alongside #cryptid, to great effect:

> The conspiracy theory crowd were completely gone from my videos — not commenting on them, not seeing them. I've just got the kind of interactions I want to have.

This algorithmic trapdoor, however, comes with its own problems. As P3, along with P4 and P17, note, queer-specific hashtags could easily be suppressed by TikTok itself. Moreover, as P6 notes, even hashtags used as part of algorithmic trapdoors have an inherent algorithmic trap attached to them — they are searchable, and therefore targetable for direct harassment:

> A lot of my Transphobic and LGBT-phobic comments were only on videos that had [community] hashtags, so I'm assuming that they search up these hashtags when they're bored just to pick on people or to go through each video on the list... I had one video where I had used the hashtag #transphobia... and it had attracted a lot of TERFs[7], which doesn't usually happen, so I'm assuming that they like to peruse that hashtag for videos.

To deal with this algorithmic trap, P6 turned to less-common and even more specific hashtags (e.g., #transgirls instead of #trans). While this strategy still has risks, P6 was able to evaluate the algorithmic doors to visibility offered to her using her folk theory of the FYP, clearly see the algorithmic traps, and find algorithmic trapdoors which secured initial distribution while mitigating unacceptably harmful potential outcomes.

4.2.2 Comment Tools as Doors with Questionable Traps

Once distribution is underway, participants widely recognized that the key algorithmic doors to further visibility via wider distribution all concern engagement, especially comments, duets, and stitches[8], the platform's three primary mechanisms for direct interaction between individual accounts. However, they also recognized the key algorithmic trap behind all these mechanisms: for the platform, all engagement appears to be treated as an indicator of who a video should be distributed to next, including explicitly transphobic and transmisogynistic engagement.

Hate, essentially, is very engaging from a computational point of view, as P11, a Jewish transfemme creator, found out during an incident where she posted a video which initially drew interactions from her regular followers, up to the point of roughly 6,000 video views, pushing the video to broader and broader audiences:

> All of a sudden, I got two or three comments from the more conservative folk, and then there was a deluge. That video is now at like 80,000 views... It definitely started with a few comments from more conservative people that got responses from my more left-leaning audience, and then it and then it really blew up over the course of two days. And then there was a 24 hour period in which it was it was being shown to neo Nazis... fights started happening in the comments, people would say some shit... That kind of engagement really seemed to spur the algorithm on.

P11 recognizes that engagement is key to visibility here, but they also recognize the algorithmic trap of undifferentiated visibility. For a Jewish trans woman's content, positive interactions appear to drive content distribution to more supportive queer people — but hateful,

---

[7] TERFs, or trans exclusionary radical feminists, are activists who oppose equal rights for transgender people, often enacting this opposition via direct harassment of trans people, disinformation campaigns, and legal challenges based on assigned sex at birth. While TERFs claim feminist motivations, their actions, goals, and rhetoric are more in line with right-wing, anti-feminist thought [40].

[8] Duets and stitches are two of TikTok's mechanisms for reposting content. Duets allow new content to be recorded alongside existing content. Stitches allow a sample of existing content to be included at the beginning of new content.

transmisogynistic, antisemitic comments from transphobes and neo-Nazis drive content distribution towards other neo-Nazis and transphobes just as well, if not better.

While P11 and most other participants are personally affected by progressively-growing torrents of algorithmically-delivered hate, they often focused more on the potential negatives of followers seeing such negative engagement, as P15 notes:

> I felt like I had a responsibility to the trans people who follow me, to the queer people who followed me, so that they didn't have to see content that was directly phobic... I want people to feel like my page is safe and fun to interact with.

P13 felt similarly, and specifically noted the importance of making sure younger followers who may be looking for trans role models or exploring their identity were not exposed to hateful commentary:

> I have so many kids that follow me, I don't want that kind of language there for the kids to see. I don't want them coming on my page, going into my comments, and seeing hurtful, awful, rude, terrible things in my comments that might make them feel bad about themselves.

Considering the concern over protecting kids especially, it may seem reasonable to argue that transfemmes have many options to mitigate this algorithmic trap: turn comments off entirely, use the comment filtering tools, or report the comments to the platform. The most obvious strategy is turning off comments– but that is the closing of a door entirely, directly preventing participants from reaching their educational and community goals by entirely rejecting the increased visibility that engagement brings. Meanwhile, the filtering and reporting systems often seem too risky or ineffective to participants, both on practical grounds and in terms of ability to work with the tool as part of a larger folk theory.

As an example, participants were broadly not fond of the keyword comment filters. Participants found the system confusing and time-consuming to use or even find, with P2 noting that the hard limit on the number of keywords imposed by the system does not even leave room for a basic list of manually-entered transphobic slurs. Participants also often found the filtering system itself to be inconsistent. For example, P3, P16, and P13 each encountered the system filtering hateful comments for them, but not for their audiences, leaving them concerned that the system was simply hiding hateful content from them while letting it stay up publicly. P13 also notes that her own comments seem to be the only comments that are consistently filtered. Moreover, many participants were worried about even turning this system on due to the potential for false positives, as P17 noted:

> If one of my followers goes like "help me the super straights[9] found me" or something like that, then I won't see it and that scares me more than the idea that I have to go and delete a comment... [the system] has just as much a chance to filter the abused as it does the abuser... I just don't want to silence people that do need to have their voices heard.

For a group of transfemme creators with explicitly educational and community-based goals, potentially losing the voices of the people they are here to help due to faulty algorithmic filtering is often an unacceptable form of trap. For many participants, these tools represent another algorithmic trap: the design decisions behind the tool appear to assume a limited vocabulary of hate speech which needs to be accounted for, and a use case where comments do not have the potential to be time-sensitive cries for help from more vulnerable people.

---

[9] A transphobic hate and trolling group which primarily exists on Reddit and TikTok, posing unwillingness to engage with transgender romantic and sexual partners as its own sexuality.

As for the option of reporting, I will more extensively cover rejection of this strategy as part of section 4.3.3. However, it does factor into FYP decision-making within the context of the expanding stages theory, as there is a perception that the very act of reporting might damage overall visibility — essentially, that reporting itself can spring an algorithmic trap, as P3 notes:

> I do not report anyone on TikTok any longer, because every time I've reported people my views have dropped off significantly... what I do tend to do is, I have a few friends that have TikTok and they don't make content, and I've had them report... but like me personally, with my account, I don't report anyone anymore, because I know that if I do, I will be suppressed for that.

4.3 Traps Not Worth Springing: Demotivational Folk Theories

In direct contrast to the extensible and adaptable expanding stages folk theory, transfeminine creators also held a number of folk theories which essentially assert algorithmic traps behind specific algorithmic doors to visibility or algorithmic doors on the platform generally. I call these **demotivational folk theories**, as they do not extensibly guide the user in evaluating options and strategies, but rather only help the user explain why the platform is behaving in an unacceptable and intractable way. These theories and the algorithmic traps they assert appear to erode participants' positive perception of TikTok's platform spirit while also motivating users to avoid the use of related visibility or safety/privacy tools.

4.3.1 The Algorithmic Paternalism Folk Theory

One particularly demotivational folk theory is that the system is actively engaged in a sort of **algorithmic paternalism**, where transfeminine creators are recognized as a distinct class of people that is likely to attract harassment — and is therefore best kept safe by keeping them separate from the rest of TikTok. Essentially, Tiktok "tries to shelter certain people" (P2) or keep them in a "safe bubble" (P16), and in doing so takes away their agency to evaluate and make decisions about their own visibility.
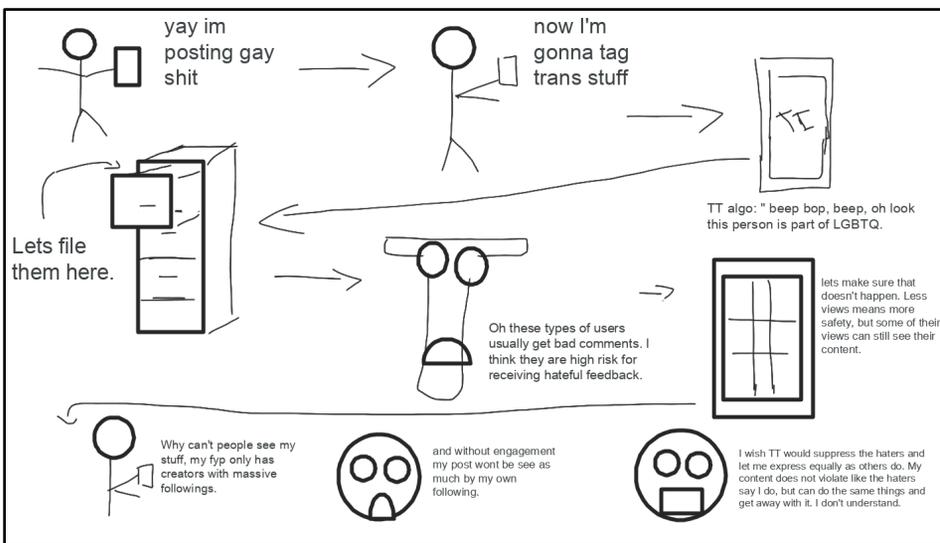

Figure 3: P8's drawing exercise showing an algorithmic paternalism-based folk theory of TikTok.

We can see a version of this theory in P8's drawing exercise (Figure 3), where posting (and especially tagging) transgender content gets creators filed as that identity specifically, which TikTok then sees as "high risk for receiving hateful feedback," limiting distribution as a protective measure. P8 does not appreciate this paternalistic attempt at protecting trans creators, noting that while this may bring a limited form of protection, it also restricts visibility and gets in the way of both her purpose for using TikTok in the first place (education) as well as tactics she uses to further that end via growing a following.

P17 also held this theory, asserting that TikTok is training their systems to avoid conflict, resulting in this paternalistic suppression. For her, even using trans hashtags on a post seemed likely to slow distribution and reduce visibility. While P17 extensively discussed her experiences getting harassed on the platform, she still sees TikTok making this potentially-protective choice on her behalf as inappropriate:

> I can take some hate if it means that I can find some more people who need to hear what I need to say... some people may need the safe space, and I'm not critiquing them, but certainly assuming that all of us need to be protected is like... a lot of us have dealt with this for a lot longer than TikTok has been around... I really don't need to be protected from a bunch of 13 year olds on the internet.

For P17 and other participants with this folk theory, the assumption that the platform will take one's visibility away for protective reasons points to another perceived algorithmic trap: to the extent that the platform understands queer and trans identities at all, it understands these identities as potential points of friction and sources of problematically controversial content. In turn, the presence of this algorithmic trap marks certain algorithmic doors to visibility, such as hashtags, and tools for maintaining that visibility safely, such as reporting, as overly risky, because they may be signals that prompt the system to act paternalistically. Moreover, this theory erodes positive perceptions of TikTok's platform spirit, because it essentially asserts that TikTok has decided that the consequences of abuse and harassment should primarily impact those targeted. As P8 notes at the end of her map:

> I wish [TikTok] would suppress the haters and let me express equally as others do.

4.3.2 The Identity Flattening Folk Theory

Another folk theory held by participants relates to identity flattening — the reduction of one's identity down to a limited set of socially acceptable attributes in environments where other attributes are discouraged or stigmatized in order to be accepted and treated well [52]. Some participants believe that TikTok, essentially, wants users to just be one, easily-categorized thing — in this case, transfeminine, and only transfeminine. P7, for example, found their healing and psychology content consistently outperformed by their trans content; P11 saw her Dungeons & Dragons content underperform compared to trans content; P15 saw her political content languish without distribution, while her trans content went wide (so long as it was not focused on the intersection between Blackness and transness).

P13, who has extensive interests in and experience with movies, television, and playwriting, notes how this creates the impression that being anything other than trans with one's visibility will fail:

> That seems to be the thing that people want to see, according to the TikTok algorithm — that seems to be the only thing that does well, is anything related to me being trans. If I'm not talking about that, it just evaporates into thin air... It sucks, to be quite honest, it's frustrating... I feel like I am not allowed to go on there and make the joke

about Riverdale[10] having terrible writing, I have to somehow make every joke a trans punch line because if I don't then no one's going to see the video.

To these transfeminine creators, the identity flattening folk theory represents a clear algorithmic trap behind the platform's algorithmic doors to visibility: the platform appears to programmatically favor content which matches its existing understanding of individual creators. This particular algorithmic trap is specifically concerning to creators, as it conflicts with the authenticity of self that most participants value, as P2 explains:

What's important to me is that I can create a safe space, and for you to feel safe around someone, you have to know things about them. And I don't want people to feel like they can't be their authentic selves in my comments sections and in my videos.

Some creators are willing to engage in trends that fit their established style (e.g., P1) and slightly emphasize certain parts of their personality (e.g., P9's habit of playing up her more extroverted tendencies). However, the desire to both show and create safety through authentically presenting themselves makes a significant collapsing of identity anathema to these participants. In turn, holding the folk theory that the platform would essentially prefer that collapsed identity has made participants think less positively about TikTok content distribution overall, damaging participants' positive perceptions of the platform's spirit.

4.3.3 The Cultural Context Folk Theory

Finally, some participants held what I call the **cultural context folk theory**, which holds that the root cause behind multiple issues creators face on the platform is a basic lack of understanding of trans culture and language. This appears to be the heart of distrust of the reporting system, as reporting comments, which seems like an obvious strategy for fighting harassment and abuse, is avoided by some participants. Participants see a pattern of rejected reports of obvious transphobia an indicator that the reporting system does not understand transmisogynistic oppressions well enough to be worth using, as P8 explains:

I can't [report] — it won't go through. I can report it, but it on what basis? On what basis is it transphobic? The algorithm doesn't... even have "tranny" in there as a slur... there's these little nuances that the algorithm should learn, and this doesn't just apply to transphobia, it applies to racism, it applies to homophobia, it applies to sexism, it applies to a lot of things — this algorithm needs to be tweaked to understand the nuances of oppression.

The cultural context folk theory asserts another algorithmic trap behind most or all algorithmic doors to visibility on the platform: reactions to transfeminine visibility that transfeminine creators find distinctly and immediately harmful do not appear to be considered at all problematic by the platform's moderation tools and/or policies. This appears to lead to an increasingly negative perception of the platform's spirit, even tipping into anger, as P10 demonstrates while directly appealing to TikTok:

Pay attention to what we're telling you... these are threats, these are direct threats, and those things, they cause people to kill themselves. They cause other people to kill us. Shut it down. Believe us when we tell you "yeah, that's a threat." When we say calling us "men in dresses" is hate speech, believe us.

Transfeminine creators have no clear information here on how the reporting system works, other than the perception that it is not even capable of understanding transmisogynistic

---

[10] A teen drama television show popular with many queer Americans.

content, resulting in a demotivational folk theory which provides no options to explore, only blame. The demotivational nature of the cultural context folk theory appears to collide with the increasingly negative perception of platform spirit, prompting some participants to avoid reporting at all — an outright abandonment of the platform's core policy enforcement feature.

5 DISCUSSION

For Gossett et al. and Pinter et al., the keys to safely and effectively pursuing trans visibility are the ability to recognize and assess traps in order to mitigate risk, and the ability to find or create trapdoors in order to escape traps entirely [28, 41]. In extending this work to the algorithmically-driven context of transfeminine visibility on social platforms, I have demonstrated that it is folk theorization that serves as the mechanism for both of these crucial abilities in algorithmic environments, while also identifying five novel folk theories of TikTok used for this very purpose. Platforms offer many *algorithmic doors* to visibility in the form of affordances and features, and all participants engaged in the process of folk theorization described in my prior work [13] in order to assess the risk level and associated costs of the *algorithmic traps* inherent in accepting any of these offers of visibility. Moreover, participants appeared to be engaging in the type of adaptation that high-level structural folk theorization enables [13] in order to find or create *algorithmic trapdoors* (e.g. the combinatory hashtag strategies discussed in section 4.2.1 and the adaptation strategies discussed by Karizat et al. and Simpson and Semaan [33, 48]). In fact, it is fair to say that all participants working from the expanding stages folk theory were deeply engaged in structural folk theorization, which focuses on understanding and using the structure of how systems make decisions, not simply what decisions they make [13]. Structural folk theorization ability, then, can be viewed as a key tool for marginalized creators as they try to safely and effectively gain visibility in algorithmically-driven environments — as they attempt to successfully navigate the *trap of algorithmic visibility*.

In turn, the ability to successfully navigate the trap of algorithmic visibility is a crucial skill for transfeminine content creators, as it directly supports these creators in achieving their goals in an algorithmic environment. As explored in section 4.1, these goals largely center around providing aid to other trans people and improving the visibility of different kinds of trans identities and lives. In the algorithmic social platform context, this can be read as deliberately combatting what Andalibi and Garcia call "algorithmic symbolic annihilation," the "emotional and representational harm caused by algorithms that perpetuate normative narratives" [4 p. 20]. Structural folk theorizing provides a safe path towards pursuing the visibility which allows creators to combat the effects of algorithmic symbolic annihilation, potentially preventing other transfeminine people from feeling erased and invalidated in their identity. This suggests that an important way to support marginalized creators with educational and community-service goals is to work towards boosting creator theorization to a structural level via algorithmic literacy interventions.

This work has also extended my folk theorization model by identifying two new classifications: *actionable* and *demotivational* folk theories. My prior model hinges on the distinction between structural and functional folk theorization, with structural theorization asserted as the overall goal for individual users, as it allows extensible, autonomous sensemaking and adaptation [13]. However, the present study suggests that structurally-based folk theories may also need to be *actionable* to both enable and motivate adaptive use of platform tools. Consider that the *algorithmic paternalism* folk theory has strong structural elements, with participants theorizing multiple inputs that could get one marked as part of a potentially-problematic group — yet the theory serves mostly to discourage any engagement with the platform's key reporting and enforcement mechanisms. While *structural* folk theorization appears to be the key user-side goal for enabling strategic sensemaking and adaptation, enabling *actionable* theories can be seen as a key platform-side goal in service of

making essential platform features theorizable enough to be positively evaluated and regularly used. Consider that all the *demotivational* folk theories found in this study had a strong component of assumed platform motivations which actively damaged positive perceptions of platform spirit, e.g., genuine attempts to protect gone terribly wrong in the algorithmic paternalism folk theory, and lack of care over trans identities in the cultural context folk theory. These findings suggest that an important way to encourage actionable theories of one's platform is to attend to perceptions of platform spirit by proactively explaining and providing transparency around mechanisms and design motivations, especially when identity is central.

Finally, while past work implies the potential for folk theories to overlap (e.g., [13, 14, 15]), this study suggests that folk theorization of platforms includes multiple, overlapping folk theories of both the whole platform and individual algorithmic components of the platform. All participants were at least theorizing about both the FYP and the content moderation systems, and all participants held multiple, overlapping folk theories. Moreover, it was only by examining folk theorization at multiple levels that this study was able to fully take the complexity of algorithmic systems and user understandings of this complexity into account. In such systems, there will always be algorithmic doors within algorithmic doors (e.g., the offer of the platform for a site for visibility as a whole and the offers of individual tools for visibility embedded within the platform); being attentive to only one level opens up the potential to misunderstand the situation. In future work, it is important for researchers to consider folk theorization across a platform environment at multiple levels, and to look for complex folk theories of platforms in aggregate in addition to folk theories of single systems.

5.1 We Don't Want This Help: Algorithmic Paternalism

Arguably the most problematic among the demotivational folk theories I found is the algorithmic paternalism theory, in which what some participants saw as actions with genuinely positive, protective motivations by the platform were still detrimental to transfeminine visibility, drawing the ire of creators. While platforms always have structural power over their users, the core of paternalism towards groups of people is the imposition of requirements or restrictions by a more powerful entity towards an entire group of people, regardless of individual variation on the issue that the paternalistic restriction purports to benevolently "solve" for said group. A benevolent motivation is a key part of this definition — this kind of imposition on a group for entirely self-serving motives such as monetization is simply exploitation [29]. While neither I nor my participants can claim direct knowledge of TikTok's motivations, participants with the algorithmic paternalism folk theory perceived TikTok's motivations as genuinely protective, restricting visibility to reduce the possibility of harassment stemming from that visibility. However, as P8 pointed out, restricting visibility protectively is still restricting visibility overall, directly contradicting the entire use case for the platform. By taking the choice to be visible away from users, for whatever reason, protective or otherwise, the platform essentially closes key algorithmic doors to visibility, one of the most important paths to progress for trans people.

It is unlikely that transfemme creators, or likely marginalized users in general, will ever accept this kind of paternalistic restriction, or even the appearance of such a restriction. On a perceptual level, my participants made it clear that this algorithmic paternalism is being read less as a positive desire to protect, and more as the inappropriate placing of blame and the onus of responsibility for harassment on the harassed, instead of the actual bad actor in the situation. On a broader level, this kind of paternalistic restriction violates HCI-relevant sets of principles such as Costanza-Chock's design justice approach, which privileges community impact over designer intent and avoids one-size-fits-all solutions [10] (which paternalism always imposes [29]). Even in the context of ubiquitous technology more generally, where creating calm, unobtrusive systems is a key goal, Spiekermann and Pallas note that users should always

ultimately have the last say on key decisions, even if it creates friction [49]. Moreover, this paternalistic restriction also directly contradicts key transfeminist principles, such as privileging individual choice over institutional limitation, and the right to have not just one's gender, but one's expression of that gender, respected and unrestricted [38].

Based on these findings, I challenge platforms to actively investigate and interrogate current systems and future designs, especially those with the intent to protect, to avoid both the appearance and reality of algorithmic paternalism. These results suggest that even if platform designers have positive intent, and this positive intent is understood by users, paternalistic designs will still result in a more negative perception of platform spirit and avoidance of key protective systems. Practically, this is also important to avoid both based on past work on negative user reactions [49] and based on modern ethical and especially design principles (e.g., [10, 38]), which require far more user autonomy than a paternalistic system can provide.

5.1.1 Paternalistic Visibility or Exploitation Via Visibility?

In addition to the algorithmic paternalism of restriction, there is may be a second form of algorithmic paternalism in play here: the paternalism of the FYP itself, which effectively uses engagement data to make visibility decisions on behalf of the user, potentially without adequate user input. That is not to say there is no user input; as demonstrated in section 4.2, there is ample input to the system that suggests algorithmic doors to visibility the creator might be interested in. However, there is no algorithmic counterpart to this deliberate opening of a door, no way for a user to assert that they are not interested in entering a door to visibility that the platform opts them into. It is debatable if this is algorithmic paternalism or simply algorithmic exploitation, as what one person perceives as paternalistic may not feel paternalistic to others [29]. Participants varied in their perceptions, with some seeing this as an overzealous platform genuinely trying to provide the most visibility for trans people (a paternalistic motivation), and others seeing this as the platform maximizing views in order to serve more advertising (an exploitative motivation). Either way, here, TikTok effectively asks creators to play what may seem like a simple game of choosing algorithmic doors, but which is in reality a deeply risky exercise in avoiding algorithmic traps for marginalized creators. TikTok provides ample opportunity to exercise agency in terms of joining the game, but very little agency around deciding when and how the creator should exit this game. Transfemme creators are not done being visible when they say they are — they're done when the FYP says they are, and that regularly and directly endangers them.

Every single participant in this study spoke of the need to carefully manage their visibility; every single participant expressed some version of the fear that content would become too visible to the wrong people, with no way for them to stop it. Platforms can improve user control over visibility by providing their own algorithmic trapdoor: a way to stop algorithmic visibility decision-making, to opt out of the platform's paternalistic and/or exploitative control over algorithmic visibility.

Platforms should consider the implementation of an **algorithmic emergency brake**, a mechanism which allows the user to remove a piece of content from either algorithmic distribution overall without losing existing views and engagement and the opportunity for future visibility (as making a video private or deleting it does now). Moreover, while a blanket stop would be a useful protective mechanism for creators which preserves their autonomy, to help preserve the potential for important visibility, platforms could go further and allow the selective restriction of visibility to only certain doors the creator has opted to open. For example, in the context of TikTok, one can envision a setting which allows caps on distribution at key thresholds, such as not leaving the follower feed, or not leaving one's "side" of TikTok.

5.2 You Just Don't Understand: Decontextualized Engagement and Judgement

The other problematic cross-cutting issue this study identified is **decontextualization**, both in the form of the cultural context folk theory, and in the comment-centric interactions described in section 4.2.2. The fact that reporting and content moderation appears to happen without much accounting for local context creates a situation where the platform may literally not be able to understand that something is or is not harassment or abuse. Consider that "you're a man" is a perfectly normal thing to say to many human beings; saying it to a transfeminine creator is hate speech — but the system provides no way to contextualize the statement with key identity information. Similarly, recall that participants reported many instances of comment interaction from the wrong side of TikTok turning into deluges of hate. Again, there is no way for the system to know that a particular comment, or other sort of engagement, or even full video view is an action taken maliciously instead of genuine, positive interaction. Also, recall the perception that there was undue and uneven enforcement of platform rules, with transfeminine creators subject to harsh enforcement and baffling content removals. Past work has established that platforms regularly misunderstand trans content as pornographic [31], and that patchy enforcement and the entire culture of certain platforms brings undue scrutiny and ineffective protection to queer women in particular [20]. As such, it seems clear that TikTok does not, in fact, understand trans identity generally or transfeminine identity in particular (as many platforms don't). In turn, this violates key design principles for transgender users such as support for erotics [31, 32], key design principles for queer users more generally such as self-determination [17], and transfeminist and feminist design principles around autonomy and self-knowledge [10, 38].

I recommend that platforms like TikTok implement ways to contextualize both interaction and reporting, allowing the user to use their own local, situational, and self-knowledge to inform either an algorithmic system or human moderator of key details — essentially, an algorithmic trapdoor which lets the user bypass lack of context. In terms of reporting, this could be as simple as an open text box which allows more specific explanation of why a report is being filed, or at least an identity-based dropdown which allows users to further contextualize reporting within categories such as "hate speech," so that the system or moderator knows to evaluate for hate speech in the specific context of transfeminine identity. In terms of engagement, platforms could consider implementing creator-side flags for interaction. For example, platforms could provide the option for a creator to mark one of their comments as "corrective," allowing them to directly respond to and counter transphobic and transmisogynistic content while signaling the system that this interaction should not be viewed as evidence that more users like the user being corrected should be shown the video. More programmatically, platforms could also use existing user actions to construct these flags on the backend, e.g., cancelling out the algorithmic input from interactions where the creator has subsequently deleted a comment or blocked the commenter.

5.3 Limitations and Future Work

While this work gives us novel insight into transfeminine creator experiences, folk theorization, and navigating trans visibility, it also has important limitations to consider when interpreting the results. While this work gives us novel insight into how transfemme content creators navigate algorithmic doors and traps, the algorithmic trap of visibility applies to all trans people [28], and it is unlikely that it is only transfemmes who wish to risk the trap while explicitly and publicly out. As such, one promising avenue for future work is comparative study within the trans population, including work on transmasculine creators and creators who identify outside of any binary conception of gender. Additionally, the trap of algorithmic visibility also applies to those who are still trying to reach 1,000 followers, the minimum cutoff for participation in this study. While the cutoff allowed this study to focus on creators with established theorization and adaptation routines, it is important for future work to examine how

creators get to this point, how their folk theories differ from those with an established following, and how we can design systems that help creators navigate the algorithmic traps involved in initially establishing wide visibility. Moreover, while the current study primarily focuses on how folk theories are strategically employed, work on newer creators would benefit from more focus on how the theories are initially formed and what information sources creators draw from. More broadly, consider that all of the demotivational folk theories discussed here could easily be reinforced by popular press articles which document TikTok's previous mishandling of identity-focused content[11]. Studying these exogenous information sources could be especially important in the context of addressing these demotivational folk theories, which appear to motivate users to avoid key platform features.

Another key limitation is that this study is entirely in the context of transfeminine identity in the United States and United Kingdom. This focus was necessary for the complete-member-researcher approach [1] used in this study, as the author is a complete-member-researcher in the context of English-language transfeminine TikTok creators. However, conceptions of gender, the nature of what is labeled as "trans," and the legal and cultural realities of existing as a visibly trans person vary from country to country and region to region [12, 35]. Combined with the existence of regional versions of platforms (e.g., Douyin instead of TikTok in China), this suggests that transfeminine creators outside of the US/UK context may have different concerns and goals and, therefore, differing folk theories and theorization habits. Future work that is based in a non-western context is key to fully understanding trans experiences with visibility in algorithmic environments.

This study also opens up an important question for future folk theories work. This study and past work (e.g., [33, 37, 48]) have found unusually high-level folk theorization occurring on TikTok. As I have previously argued, high-level, especially structural-level, folk theorization is a key to algorithmic literacy and user autonomy [13]. As such, future work that investigates why TikTok appears to be an unusually theorizable platform could help further those goals.

5.3.1 Race and Racialized Identities

The most important limitation of this study concerns race and racialized identities. While this study does include creators of color, the sample is majority white and this is reflected in the results and interpretation. Alcoff argues that the two most fundamental, and therefore two most practically impactful, aspects of identity are race and gender, as unlike sexuality or economic class, they are always and immediately visible [3]. As the central focus of this work is transfeminine identity, visibility, and adaptation in spite of transmisogyny, I have largely focused on gender here. However, there is good reason to expect that focusing specifically on the experiences of transfeminine people of color would deepen and further contextualize these results in important ways. While participants of color in this study had overall similar experiences to white participants, creators such as P5 and P15 specifically noted race as a factor in a way that white creators did not. For example, P5 was more attentive to the realities of race and visual media, commenting on the perceived importance of lighting that makes people of color look lighter-skinned. Similarly, P15's interview identified key differences in how she conceptualizes the algorithmic community folk theory as limited and unidimensional as compared to white creator perceptions. Both examples suggest that transfeminine creators of color have a different, more wary, and more skeptical approach to evaluating algorithmic traps. This reflects an inherent awareness of the core basis of the traps concept: externally imposed bounds on acceptable trans visibility often do not allow for trans visibility around race [28]. It is reasonable to assume that operating with this more skeptical approach to evaluation would have impacts on perceptions of platform spirit, where lines are drawn between acceptable and

---

[11] https://slate.com/technology/2019/12/tiktok-disabled-users-videos-suppressed.html

unacceptable risk when considering algorithmic doors to visibility, and perhaps even perceptions of how actionable or demotivational a folk theory is.

Moreover, there are other key differences to consider around race and visibility. The identity and behavioral expectations placed on trans people vary on cultural and racial lines, and in a US context tend to be understood relative to "androcentric, heterocentric, middle-class whitenormative cultural narratives" [12 p. 63]. This changes how trans people of color think about visibility. For example, in comparison to white transfemmes, Black transfemmes must contextualize their own visibility while contending with more concerns about the difference between performance and exploitation, a public lack of awareness of Black trans history, and very little concrete sense of a larger trans community [27]. This is in addition to the general stress of being visibly transgender while significantly more marginalized and at risk for harassment and physical harm than white peers[12].

These race-based differences around identity create a heightened need for trans creators of color to learn (and learn how to either resist or usefully embrace) dominant cultural narratives of visibility [12]. At the same time, transfeminine creators of color play a particularly crucial role when they pursue the types of goals covered in this study, as there are large, harmful disparities in health information and access for transgender people of color [34]. In combination, this suggests that the goal of better supporting transfeminine creators of color is crucial for peer education among creators and, in turn, providing critical information to the people these creators serve. As such, future work in this area which focuses squarely on race and the adaptive experiences of Black, Indigenous, Latinx, and Asian transfeminine creators is crucial.

6   CONCLUSION

In this paper, I have contributed a new understanding of how transfeminine content creators navigate the algorithmic trap of visibility, five novel folk theories employed in this navigation, and extensions of both folk theorization and the trap of visibility lens. Moreover, by working within my own community as a complete-member-researcher, I have illustrated how one community-minded subset of a marginalized group navigates significant algorithmic risk to provide education, aid, and social support services centering around a shared identity. Learning from this example, I have provided directions on how we as researchers and designers can better support those who are willing to brave the trap of algorithmic visibility for the benefit of their communities. However, is crucial to note that transfeminine people do not stand alone; transfeminine liberation is inherently tied to the liberation of other marginalized people [38]. It is my hope that this work represents a first step towards better supporting and building upon existing folk theorization as a tool for anyone attempting to brave algorithmic traps in service of their own marginalized community.

REFERENCES

1.   Patricia A Adler and Peter Adler. 1987. *Membership Roles in Field Research*. Sage, Thousand Oaks, California.
2.   Alex A Ahmed. 2018. Trans Competent Interaction Design: A Qualitative Study on Voice, Identity, and Technology. *Interacting with Computers*, 30, 1: 53-71.

---

[12] https://www.hrc.org/resources/violence-against-the-trans-and-gender-non-conforming-community-in-2020

3. Linda Martín Alcoff. 2005. *Visible identities: Race, gender, and the self.* Oxford University Press
4. Nazanin Andalibi and Patricia Garcia. 2021. Sensemaking and coping after pregnancy loss: the seeking and disruption of emotional validation online. *Proceedings of the ACM on Human-Computer Interaction*, 5, CSCW1: 1-32.
5. Monica Anderson and Emily A Vogels. 2020. *Young women often face sexual harassment online–including on dating sites and apps.* Pew Research Center, Washington, DC.
6. Matthew Carrasco and Andruid Kerne. 2018. Queer visibility: Supporting LGBT+ selective visibility on social media. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1-12.
7. Kathy Charmaz. 2006. *Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis.* Sage, Thousand Oaks, CA.
8. Kathy Charmaz and Liska Belgrave. 2012. Qualitative interviewing and grounded theory analysis. In *The SAGE handbook of interview research: The complexity of the craft* (2), Jaber F. Gubrium, James A. Holstein, Amir B. Marvasti and Karyn D. McKinney (ed.). Sage, Thousand Oaks, CA, 347-365.
9. John Cheney-Lippold. 2011. A new algorithmic identity: Soft biopolitics and the modulation of control. *Theory, Culture & Society*, 28, 6: 164-181.
10. Sasha Costanza-Chock. 2020. *Design justice: Community-led practices to build the worlds we need.* The MIT Press, Cambridge, MA.
11. Kelley Cotter. 2018. Playing the visibility game: How digital influencers and algorithms negotiate influence on Instagram. *New Media & Society*, 21, 4: 895–913.
12. Kylan Mattias De Vries. 2012. Intersectional identities and conceptions of the self: The experience of transgender people. *Symbolic Interaction*, 35, 1: 49-67.
13. Michael Ann DeVito. 2021. Adaptive Folk Theorization as a Path to Algorithmic Literacy on Changing Platforms. *Proceedings of the ACM on Human-Computer Interaction*, 5, CSCW2: 339.
14. Michael Ann DeVito, Jeremy Birnholtz, Jeffrey T. Hancock, Megan French, and Sunny Liu. 2018. How People Form Folk Theories of Social Media Feeds and What It Means for How We Study Self-Presentation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1-12.
15. Michael Ann DeVito, Darren Gergle, and Jeremy Birnholtz. 2017. "Algorithms ruin everything": #RIPTwitter, Folk Theories, and Resistance to Algorithmic Change in Social Media. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 3163-3174.
16. Michael Ann 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.
17. Michael Ann DeVito, Ashley Marie Walker, and Julia R Fernandez. 2021. Values (Mis) alignment: Exploring Tensions Between Platform and LGBTQ+ Community Design Values. *Proceedings of the ACM on Human-Computer Interaction*, 5, CSCW1: 1-27.
18. Maeve Duggan. 2017. *Online Harassment 2017.* Pew Research Center, Washington, DC.
19. Stefanie Duguay. 2016. "He has a way gayer Facebook than I do": Investigating sexual identity disclosure and context collapse on a social networking site. *New media & society*, 18, 6: 891-907.
20. Stefanie Duguay, Jean Burgess, and Nicolas Suzor. 2020. Queer women's experiences of patchwork platform governance on Tinder, Instagram, and Vine. *Convergence*, 26, 2: 237-252.
21. Sonya Corbin Dwyer and Jennifer L. Buckle. 2009. The Space Between: On Being an Insider-Outsider in Qualitative Research. *International Journal of Qualitative Methods*, 8, 1: 54-63.
22. Motahhare Eslami, Karrie Karahalios, Christian Sandvig, Kristen Vaccaro, Aimee Rickman, Kevin Hamilton, and Alex Kirlik. 2016. First I "like" it, then I hide it: Folk Theories of Social Feeds. In *Proceedings of the 34rd Annual SIGCHI Conference on Human Factors in Computing Systems*, 2371-2382.
23. Motahhare Eslami, Aimee Rickman, Kristen Vaccaro, Amirhossein Aleyasen, Andy Vuong, Karrie Karahalios, Kevin Hamilton, and Christian Sandvig. 2015. "I always assumed that I wasn't really that close to [her]": Reasoning about invisible algorithms in the news feed. In *Proceedings of the 33rd Annual SIGCHI Conference on Human Factors in Computing Systems*, 153-162.
24. 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.
25. Tarleton Gillespie. 2014. The Relevance of Algorithms. In *Media Technologies: Essays on Communication, Materiality, and Society*, Tarleton Gillespie, Pablo Boczkowski and Kirsten Foot (ed.). MIT Press, Cambridge, MA, 167-193.
26. Tarleton Gillespie. 2018. *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media.* Yale University Press
27. Che Gossett and Juliana Huxtable. 2017. Existing at the World: Blackness at the Edge of Trans Visibility. In *Trap door: Trans cultural production and the politics of visibility*, Reina Gossett, Eric A Stanley and Johanna Burton (ed.). MIT Press, Cambridge, MA, 23-38.
28. Reina Gossett, Eric A Stanley, and Johanna Burton. 2017. Known Unknowns: An Introduction to Trap Door. In *Trap door: Trans cultural production and the politics of visibility*, Reina Gossett, Eric A Stanley and Johanna Burton (ed.). MIT Press, Cambridge, MA, xv-xxvi.
29. Kalle Grill. 2018. Paternalism by and towards groups. In *The Routledge Handbook of the Philosophy of Paternalism*, (ed.). Routledge, 46-58.

30. Oliver Haimson. 2018. Social media as social transition machinery. *Proceedings of the ACM on Human-Computer Interaction*, 2, CSCW: 1-21.
31. 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.
32. 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 the Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1-13.
33. Nadia Karizat, Daniel Delmonaco, Motahhare Eslami, and Nazanin Andalibi. 2021. Algorithmic Folk Theories and Identity: How TikTok Users Co-Produce Knowledge of Identity and Engage in Algorithmic Resistance. *Proceedings of the ACM on Human-Computer Interaction*, 5, CSCW2: 305.
34. Shanna K Kattari, N Eugene Walls, Darren L Whitfield, and Lisa Langenderfer-Magruder. 2015. Racial and ethnic differences in experiences of discrimination in accessing health services among transgender people in the United States. *International Journal of Transgenderism*, 16, 2: 68-79.
35. Mary E Kite, LaCount J Togans, and Tollie J Schultz. 2019. Stability or change? A cross-cultural look at attitudes toward sexual and gender identity minorities. In *Cross-Cultural Psychology: Contemporary Themes and Perspectives*, Kenneth D Keith (ed.). 427-448.
36. Erin Klawitter and Eszter Hargittai. 2018. "It's Like Learning a Whole Other Language:" The Role of Algorithmic Skills in the Curation of Creative Goods. *International Journal of Communication*, 1221.
37. Daniel Klug, Yiluo Qin, Morgan Evans, and Geoff Kaufman. 2021. Trick and Please. A Mixed-Method Study On User Assumptions About the TikTok Algorithm. In *Proceedings of the 13th ACM Web Science Conference*, 1-9.
38. Emi Koyama. 2003. The transfeminist manifesto. In *Catching a Wave: Reclaiming Feminism for the 21st Century*, Rory Dicker and Alison Piepmeier (ed.). Northeastern University Press, Boston, MA, 244-259.
39. Marjan Nadim and Audun Fladmoe. 2021. Silencing women? Gender and online harassment. *Social Science Computer Review*, 39, 2: 245-258.
40. Ruth Pearce, Sonja Erikainen, and Ben Vincent. 2020. TERF wars: An introduction. *The Sociological Review*, 68, 4: 677-698.
41. Anthony T Pinter, Morgan Klaus Scheuerman, and Jed R Brubaker. 2021. Entering Doors, Evading Traps: Benefits and Risks of Visibility During Transgender Coming Outs. *Proceedings of the ACM on Human-Computer Interaction*, 4, CSCW3: 1-27.
42. Jonathan L Poquiz, Claire A Coyne, Robert Garofalo, and Diane Chen. 2021. Comparison of gender minority stress and resilience among transmasculine, transfeminine, and nonbinary adolescents and young adults. *Journal of Adolescent Health*, 68, 3: 615-618.
43. Emilee Rader and Rebecca Gray. 2015. Understanding User Beliefs About Algorithmic Curation in the Facebook News Feed. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 173-182.
44. 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.
45. Kristen Schilt and Matthew Wiswall. 2008. Before and after: Gender transitions, human capital, and workplace experiences. *The BE Journal of Economic Analysis & Policy*, 8, 1.
46. Julia Serano. 2007. *Whipping girl: A transsexual woman on sexism and the scapegoating of femininity*. Hachette UK
47. Perrē L Shelton. 2018. Reconsidering femme identity: On centering trans* counterculture and conceptualizing trans* femme theory. *Journal of Black Sexuality and Relationships*, 5, 1: 21-41.
48. Ellen Simpson and Bryan Semaan. 2021. For You, or For" You"? Everyday LGBTQ+ Encounters with TikTok. *Proceedings of the ACM on Human-Computer Interaction*, 4, CSCW3: 1-34.
49. Sarah Spiekermann and Frank Pallas. 2006. Technology paternalism–wider implications of ubiquitous computing. *Poiesis & praxis*, 4, 1: 6-18.
50. Susan Stryker. 2013. *(De) subjugated knowledges: An introduction to transgender studies*. Routledge
51. Jirassaya Uttarapong, Jie Cai, and Donghee Yvette Wohn. 2021. Harassment Experiences of Women and LGBTQ Live Streamers and How They Handled Negativity. In *Proceedings of the ACM International Conference on Interactive Media Experiences*, 7-19.
52. Ashley Marie Walker and Michael Ann DeVito. 2020. "'More gay' fits in better": Intracommunity Power Dynamics and Harms in Online LGBTQ+ Spaces. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1-15.
53. Bradley M Weisz, Diane M Quinn, and Michelle K Williams. 2016. Out and healthy: Being more "out" about a concealable stigmatized identity may boost the health benefits of social support. *Journal of health psychology*, 21, 12: 2934-2943.