Exploring the Utility Versus Intrusiveness of Dynamic Audience Selection on Facebook

SINDHU KIRANMAI ERNALA, Georgia Institute of Technology, USA
STEPHANIE S. YANG, Georgia Institute of Technology, USA
YUXI WU, Georgia Institute of Technology, USA
RACHEL CHEN, Georgia Institute of Technology, USA
KRISTEN WELLS, Georgia Institute of Technology, USA
SAUVIK DAS, Georgia Institute of Technology, USA

In contrast to existing, static audience controls that map poorly onto users' ideal audiences on social networking sites, dynamic audience selection (DAS) controls can make intelligent inferences to help users select their ideal audience given context and content. But does this potential utility outweigh its potential intrusiveness? We surveyed 250 participants to model users' ideal versus their chosen audiences with static controls and found a significant misalignment, suggesting that DAS might provide utility. We then designed a sensitizing prototype that allowed users to select audiences based on personal attributes, content, or context constraints. We evaluated DAS vis-a-vis Facebook's existing audience selection controls through a counterbalanced summative evaluation. We found that DAS's expressiveness, customizability, and scalability made participants feel more confident about the content they shared on Facebook. However, low transparency, distrust in algorithmic inferences, and the emergence of privacy-violating side channels made participants find the prototype unreliable or intrusive. We discuss factors that affected this trade-off between DAS's utility and intrusiveness and synthesize design implications for future audience selection tools.

CCS Concepts: • **Human-centered computing → Empirical studies in collaborative and social computing**; **Empirical studies in HCI**; • **Information systems → Social networking sites**; • **Security and privacy → Social aspects of security and privacy**.

Additional Key Words and Phrases: social media, audience controls, privacy, Facebook, dynamic audience selection, Human-AI Interaction, Privacy through Design

ACM Reference Format:
Sindhu Kiranmai Ernala, Stephanie S. Yang, Yuxi Wu, Rachel Chen, Kristen Wells, and Sauvik Das. 2021. Exploring the Utility Versus Intrusiveness of Dynamic Audience Selection on Facebook. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 342 (October 2021), 30 pages. https://doi.org/10.1145/3476083

1 INTRODUCTION

"Posting to a social network site is like speaking to an audience from behind a curtain. The audience remains invisible to the user: while the invitation list is known, the final attendance is not." [8]. Knowledge of one's audience influences whether, what and how one communicates [25]. Indeed,

Authors' addresses: Sindhu Kiranmai Ernala, sernala3@gatech.edu, Georgia Institute of Technology, USA; Stephanie S. Yang, syang479@gatech.edu, Georgia Institute of Technology, USA; Yuxi Wu, yuxiwu@gatech.edu, Georgia Institute of Technology, USA; Rachel Chen, rwchen@gatech.edu, Georgia Institute of Technology, USA; Kristen Wells, kjw3201@gatech.edu, Georgia Institute of Technology, USA; Sauvik Das, sauvik@gatech.edu, Georgia Institute of Technology, USA.

2573-0142/2021/10-ART342 $15.00
https://doi.org/10.1145/3476083

one may speak casually and freely to a spouse, but formally and only about specific topics with a colleague. Unsurprisingly, then, audience selection controls are of critical importance to social media platforms like Facebook, where people may communicate with many audiences at once.

Despite the critical importance of audience in posting decisions, audience selection controls have remained largely unchanged for over a decade: people select between pre-configured static lists of contacts. On Facebook and Instagram, users typically choose between all direct connections (i.e., friends and followers) or "public"; otherwise, they create manual lists such as "close friends" or "acquaintances" to support finer-grained audience selection. Platforms like Google+ aimed to move past the 'all-or-nothing' models for sharing on social networking sites by enabling users to selectively share (and exclude) content with a more varied set of pre-configured 'Circles' of people [29] (e.g., "family", "college friends"). More recently, Twitter added a feature allowing users to limit who replies to their post (either to those who follow the user or those tagged in the post[1].) However, these static approaches to audience selection have well noted shortcomings [45].

Indeed, prior work suggests that owing partially to the limited expressiveness of static audience selection controls, social media users navigate nuanced considerations and trade-offs related to their audience that affects whether, how, and what they end up posting [38]. The lack of awareness of one's audience [24], visibility into algorithmic curation of content [21], and misalignment between imagined and actual audiences [8] have all led to well-documented privacy violations [3, 37, 45, 53]. Consequently, users adopt a variety of boundary regulation strategies to navigate context collapse in a network composed of multiple, overlapping social circles folded into one [23, 37]. One notable strategy is only posting content for the lowest common denominator [28] — i.e., only sharing content that it is appropriate for all overlapping social circles. Another strategy is self-censorship [37, 60]. Prior work found that people censor more when their audience is hard to define, and when the relevance or topicality of the content they would share is narrower [17]. Users also experience "regret" for sharing content that might be inappropriate to some parts of their audience [58]. In short, difficulties with audience specification on social network services can lead to self-censorship, regret, and privacy violations, in turn limiting the potential benefits people may draw from self disclosure and posting to an audience on social networking platforms [14, 19].

Owing to these challenges with static audience selection controls, prior work suggests that there is potential utility in designing dynamic audience selection controls that better align users' actual and imagined audiences — i.e., by allowing users to constrain their audience based on *AI-inferrable* attributes that users can easily specify (e.g., content shared, time of day, relationships an individual has with others) [45]. Imagine being able to restrict one's audience to, e.g., "only friends who like basketball." Through this dynamic, post-specific audience specification, users might feel like there is greater alignment between their intended and actual audience for the content they share on social media [31, 45]. These sorts of intelligent audience controls are not far-fetched: ad publishers leverage AI-powered methods to target highly granular customer segments with behavioral advertisements [54]. Furthermore, recent work has highlighted the feasibility of using intelligent systems to support people's social media posting decision-making behavior [30].

Intelligent audience selection controls may unlock more utility — by bridging together actual and imagined audiences — but that does not necessarily mean they are a good idea. Employing AI to facilitate the construction of intelligent audience selection tools requires an algorithm to make potentially invasive inferences about one's contacts, e.g., who likes basketball, or who leans more politically conservative. This trade-off — between the utility and intrusiveness of dynamic audience selection controls — remains unstudied from the end users' perspective and must be understood in

[1]https://blog.twitter.com/en_us/topics/product/2020/testing-new-conversation-settings.html

order to determine if such controls *ought* to be implemented. Accordingly, in this paper, we explore dynamic audience selection (DAS) on Facebook and ask the following research questions:

RQ 1. What are the differences between users' ideal audience and their chosen audience for posts shared on Facebook?

RQ 2.1. What is the perceived utility of dynamic audience selection on Facebook?

RQ 2.2. What is the perceived intrusiveness of dynamic audience selection on Facebook?

RQ 2.3. How do users perceive the trade-offs between utility and intrusiveness of dynamic audience selection mechanisms on Facebook?

We present a two-phased study to answer these questions. First, to answer RQ1, we fielded a survey to empirically model individuals' ideal vs. their chosen audiences on recently shared Facebook posts. Understanding this chasm allowed us to operationalize the potential "utility" of DAS. Drawing on insights from this survey, we identified categories of audience constraints that participants used to specify their ideal audience. Then, to answer RQ2, we designed and evaluated a high-fidelity prototype of DAS through a counterbalanced summative evaluation. Given that our core research objective was to explore whether DAS *ought* to be implemented, we designed our prototype of DAS as a sensitizing concept — i.e., an artifact that is meant to sensitize other designers to many new possibilities beyond the specific artifact that was created [11].

We found that users' descriptions of their intended audience are highly subjective and personal, presenting significant gaps between their intentions and the existing affordances of static audience selection controls provided by Facebook. Our findings from the user study suggest that the expressiveness, reflexivity and scalability provided by dynamic audience selection helped users feel empowered to post content that they otherwise would not share due to an expected misalignment between their desired and actual audience. However, due to low transparency, distrust in algorithmic inferences, and the emergence of privacy-violating side channels, some participants perceived the prototype as unreliable or intrusive. In short, we contribute the first insights into how users perceive the trade-offs between utility and intrusiveness of dynamic audience selection on Facebook. Based on these insights, we synthesize design implications for future audience selection tools on social media.

2 BACKGROUND

2.1 Tensions in audience selection on social networking sites.

Extensive literature in computer-mediated communication and HCI discusses the tensions in audience selection on social networking sites [28, 35, 37, 53, 57]. End-users have a nebulous, imaginary conception of their audience and often underestimate their size [8, 26], indicating a lack of knowledge of their audience on social networking sites. Although users might have a selective audience in mind while sharing content [31, 40], they are largely unaware of audience management tools on platforms or do not utilize them [44, 45]. Alternately, users adopt a variety of boundary regulating strategies to maintain consistency of presentation across multiple social contexts [3, 23, 29, 38, 60], to avoid context collapse [37], not to overwhelm audience information streams [23], and to avoid criticism and disagreement [27]. A common theme across the boundary regulation strategies is a focus on what is shared (by modifying or censoring content) than with whom the content is shared (adopting audience selection controls provided by the platform) [12, 44]. For instance, users share content that would be reasonable across multiple social contexts – for the "lowest common denominator" [28, 37], or they think about posting but do not end up doing so [27, 45]. Studies on last minute self censorship on social networking sites found that people censor more when their audience is harder to define or when the topical relevance is narrower [17] or because the available audience controls were too static [45] — instead, they wanted dynamic,

contextually-dependent audience controls (e.g., people who like comic books). In other cases, prior work found that users create a completely new, second profile to deliberately partition their audience to a small subset and share a more authentic performance with such an audience [48, 61].

Thus, providing users with more awareness and control over who sees their content remains a longstanding challenge for social networking sites. Increasingly, the relevance of this challenge has become more prominent as social networking platforms become more and more pervasive within diverse social settings (workplaces [13, 44], schools [41, 49], etc.) We extend this body of work by exploring the design considerations, utility, and intrusiveness of a dynamic audience selection tool.

2.2 Audience selection controls on social networking sites.

Currently, social networking platforms articulate social connections with broad terms such as "friends," "friends of friends," "public," "followers," etc. While this terminology is easy for users to understand, prior work suggests that more granular groupings can help mitigate unintended consequences due to a mismatch between users intended and actual audience [31, 45, 58]. However, a more granular grouping of users' audience might not be straightforward. In fact, Litt and Hargittai found that although users often interact with a large, diverse audience on social networking sites, they coped with audience challenges by conceiving an abstract imagined audience [35]. These imagined audiences were not stable and instead fluctuated among different types [34, 35, 37]. There is often a misalignment between users' imagined audience and those who end up seeing their posts [8]. Thus, improving current audience selection mechanisms to capture more granular groupings of audiences requires a shift from the static, predefined controls to dynamic, customizable controls.

Researchers and platforms have innovated alternative controls for audience selection on social networking sites. One set of approaches are based on the design of access control lists. Tootoonchian et al. proposed Lockr, a social access control lists based approach for better privacy on Facebook [52]. Watson et al. proposed AudienceView, which provides different views of one's Facebook profile from different audiences [59]. Researchers explored group-based controls [31] or lists that are widely adopted by platforms like Facebook, Instagram, and Twitter today. Egelman et al. designed an interface based on Venn diagrams for specifying access control settings in Facebook [18]. Other approaches to audience selection controls include Google+ use of "circles" to selectively share or exclude content from individuals [29], however these circles required users to manually sort individuals and apply the specification prior to posting content.

Common across existing approaches is that audience selection controls are content agnostic, static, requiring manual curation, and maintenance. Additionally, current controls are insensitive to content and context constraints that might change post-by-post. In this paper, we draw on suggestions from this literature: that better controls might help mitigate last minute self-censorship, and improve privacy perceptions [31, 45, 58] on social networking sites. We build upon this body of work with the design of DAS, by incorporating granular audience selection, dynamic constraints and post-by-post audience selection into our design of DAS.

2.2.1 Feasibility of dynamic, intelligent audience selection tools. Recent advances in AI-powered user modeling suggest that intelligent audience selection / privacy configuration tools are not far-fetched. For instance, Kurtan and Yolum [32] present a user agent that makes privacy recommendations for image sharing on online social networking sites. Using previous sharing behaviors, the user agent automatically predicts the privacy setting for a new image that a user uploads. Similarly, Misra and Such [39] developed a personal assistant that makes personalized recommendations for access control in social media, based on social network relationships and the post's content. Focusing on access control policies in social networking sites, researchers have built recommendation systems

using techniques such as association rule mining [4] and clustering analysis [15] based on past sharing behaviors, social network relationships, and tie strength. A common challenge across these approaches is accounting for the context in which a post or photo is being shared on social media. Consequently, researchers have made efforts to capture context from the content itself (such as scene context obtained from images [51]), from semantic knowledge (leveraging ontology graphs and semantic similarities between dyads' sharing behaviors [33]), and from group membership [66] to improve the prediction of privacy recommendations. Relatedly, Amershi et al. developed ReGroup, an interactive machine learning system to create on-demand groups in social networks [5]. We draw inspiration from this body of literature to discuss the feasibility of features designed for our DAS prototype.

2.3 Challenges in designing dynamic, intelligent audience selection tools.

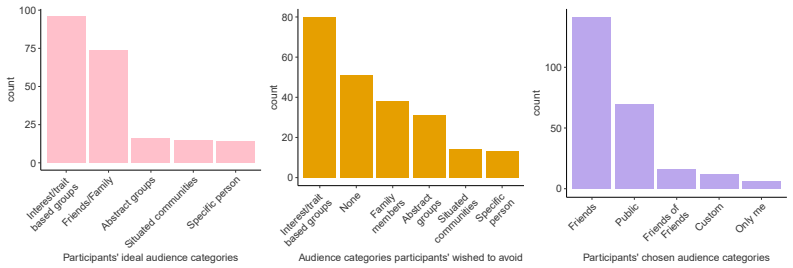
Although prior work has identified shortcomings with static audience selection controls and has prescribed the need for more dynamic audience selection mechanisms on social networking sites [17, 45], the design of audience controls has remained largely unchanged for over a decade. Why is it hard to design novel, dynamic audience constraints for social networking sites?

First, AI technologies come with error and uncertainty. As noted by Yang et al., attributes like "capability uncertainty" i.e., the uncertainties surrounding what DAS can do and how well it performs audience categorization present challenges to Human-AI (HAI) interaction design [64]. This capability uncertainty could reduce the perceived utility of dynamic audience selection controls. Second, intelligent, dynamic audience selection tools that can make finer grained predictions to categorize one's social connections might be perceived as intrusive. For example, Wang et al. found that nudges designed to make users consider the content and audience of their online disclosures were perceived as overly intrusive [58]. Similarly, in the context of online behavioral advertising, non-technical users found the idea of being specifically targeted by advertisements based on inferred interests and attributes "smart, useful, scary, and creepy at the same time" [55]. The introduction of new personalized analytics features/products is also linked to feelings of intrusiveness or "creepy" by end users [50]. This predicament between personalization and privacy mirrors the "privacy personalization paradox" where, in order to reap the potential privacy benefits of personalization, users must first share private information with a platform or service [16, 62]. In other words, dynamic audience selection controls could be considered intrusive.

How can we evaluate the balance between the potential utility of dynamic audience selection as suggested by Social Computing literature, with the potential intrusiveness and design challenges highlighted by the usable privacy and HAI literature? We address this challenge in our work by designing and evaluating a sensitizing prototype, DAS. Instead of prescribing specific attributes of DAS, we designed the prototype as a general direction along which DAS might exist on Facebook [10]. In doing so, we can evaluate DAS's utility, intrusiveness and the trade-offs therein. Our methodology allows us to answer if social networking platforms like Facebook *ought* to implement something like DAS, and opens up the design space for future innovations into intelligent audience selection tools.

3 RQ1. MODELING USERS' IDEAL AUDIENCE VERSUS THEIR CHOSEN AUDIENCE

Per RQ1, we first modeled the differences between users' ideal audience — expressed in their own terms — and their chosen audiences for specific posts shared on Facebook. In contrast to a perceived audience (who ends up seeing one's post) or imagined audience (who they had in mind while posting) in this study, we explore people's ideal audience, i.e., the people with whom users wished to share — or not share — their post. Based on the differences between their ideal and chosen



(a) (b) (c)

Fig. 1. Descriptive statistics on participants' ideal audience categories i.e. (a) with whom they wished to share, (b) with whom they wished to avoid sharing and (c) their chosen audience categories using existing Facebook controls for a most recent post.

audiences, we identify audience categories that could be supported by a dynamic audience selection tool.

3.1 Method

We fielded a survey in August 2020 and recruited 250 participants from Prolific, an on-demand participant recruitment platform. Participants were at least 18 years old, English speaking and had an active Facebook profile through which they shared content in the three months immediately preceding their participation. The full questionnaire is attached in the supplementary materials for review.

3.1.1 Survey content.

Social media posting history and preferences: Participants answered questions about their most recent post on Facebook. We asked them to report the nature of the post (personal updates, politics, entertainment, etc.) and who they shared it with using existing controls such as "only me," "friends," "friends of friends," "public," or "custom" on Facebook. We also asked them to describe their ideal audience with whom they would and would not want to share the post and why. Participants responded to questions about their audience preferences in open-ended free-form text. As an optional question, if they felt comfortable sharing, we requested participants to provide the content of this post to better contextualize their audience preferences.

IUIPC: To capture participants' general attitudes towards online privacy, we had participants answer 10 items from the Internet User Internet Privacy Concern Scale [36], eliciting their privacy practices related to control, awareness, and data collection.

Demographic covariates: Participants answered questions about demographics (gender, age, education), internet use, media exposure, and their experience with privacy violations.

3.1.2 Ethics and Compensation.

Our study was approved by an IRB. We compensated participants $2.54 for a questionnaire that took them, on average, 9 minutes to complete (equating to an average hourly wage of: $16.9).

3.1.3 Analysis.

Two researchers separately coded participants' open-ended responses on audience preferences. The researchers first identified descriptors of participants' ideal audience and then generated codes for the common audience categories in the data using thematic analysis [22]. Then, the same two researchers met to resolve disagreements and collaboratively develop a single codebook for audience categories. For the closed-ended responses to the IUIPC scale, we calculated descriptive statistics.

3.2 Results

Overall, 32% of our participants were female, 61% were between 18 and 24 years old, and 87% had more than 7 years of internet experience. In terms of educational background, 29% had some college experience but no degree, 26% had a bachelor's degree and 23% had a high school graduate degree.

3.2.1 Chosen audience using existing Facebook controls. The majority of participants (57%) chose "friends", the default audience control on Facebook, as their audience for their most recent post. Only 12 participants (4.8%) used custom controls to choose a more granular audience.

3.2.2 Ideal audience. Participants' responses revealed more nuanced preferences for their ideal audience in contrast to the audience they chose through Facebook's static audience controls. We identified six audience categories based on participants' descriptions of their ideal audience—i.e., groups of individuals with whom they wished to share or not share the post: 1) interest-/trait-based groups, 2) friends and family, 3) situated communities, 4) everyone, 5) a specific person, and 6) abstract groups. Overall, we found that participants' descriptions of their ideal audience were highly personal (e.g., my enemy, that boy I played a video game with), and did not map well onto the descriptions articulated by the platform (such as only me, friends, friends of friends, and public.)

Furthermore, a significant number of participants (34.7%) expressed multiple and overlapping ideal audiences. For instance, one participant described their ideal audience as "*group of runners, people who like sports, friends who share this passion*". Other participants had additional situational constraints for their ideal audience, such as "*my ideal audience for the post I shared were college juniors majoring in economics - as I was selling/advertising my tutoring services.*" In this case, participants' ideal audience controls were parallel to micro-targeted advertising: i.e., picking a highly selective audience among one's broader set of friends for the purposes of prompting action (in this case, purchasing tutoring services).

Interest-/trait-based groups: When participants described their ideal audience, they most often articulated an interest or trait-based grouping of people (39%). For example, people listed "*friends who watch similar TV shows*," "*those with a particular sense of humor,*" "*people feeling lonely and lost,*" etc. While descriptions of this audience category often involved more objective interest/traits such as "*anime fans*", participants also listed personal, subjective groupings like, "*people who cared about my hobbies.*" Another factor distinguishing people within this category was whether the ideal audience was a subset or a superset of one's friends on Facebook. In some cases, participants described groups of people who were a subset of their friends on Facebook, like "*friends with the same music taste.*" In other cases, participants identified broader interest/trait based groups that would include those outside their friends on Facebook — e.g., "*netflix users,*" "*people who want to make money.*" Interest and trait-based groups are not supported by Facebook's existing static audience controls, though users may be able to join interest or trait-specific groups if they regularly share content with that ideal audience in mind.

Friends and family: Another common category of audience included participants' friends and family members (30%). Within this category, participants occasionally included additional constraints like location or demographics (e.g. "*family in Greece,*" "*elder family members.*"). Facebook's existing audience controls allows for sharing with friends in general, but users would have to manually create a custom list for family members and other constrained subsets of friends.

Situated communities: Participants' ideal audience sometimes included situated communities to which they belonged, such as workplaces, college campuses, and neighborhoods (6.1%). For instance, people listed "*people who live where I do,*" "*fellow college students,*" and "*coworkers.*" As with other categories, communities were occasionally described with additional constraints such as "*people in my country who can actually make a change.*" Groupings based on situated communities

are not supported by Facebook's existing audience controls, though users might join Facebook groups for the specific community and post there.

Abstract groups: Participants sometimes did not have any constraints or specific target audience in mind and preferred to share content with everyone (6.5%). This category includes listing "*everyone*," "*public*," and "*as many people*" as the ideal audience for Facebook posts. Abstract descriptions such as "*my contacts*" were also labeled under this category. Facebook's current audience controls allow users to select "public" settings that might align with users' ideal audience when they do not have any specific constraints. However, some of the abstract descriptions like "*my contacts*" are only roughly approximated through the "friends" audience selection control.

Specific person: Finally, a few participants thought about one specific individual as their ideal audience for a Facebook post (5.7%). This category includes specific descriptions such as "*dad*," "*my cousin,*" as well as vague ones like ' '*my friend who got an internal joke*" or "*this boy with whom I'm playing a game.*" Targeting a specific person is currently supported by Facebook's audience controls through the "Specific friends" control.

We find that the above categories also captured groups with whom participants did not want to share their posts. Interest- and trait-based groups were the most common groups (32.4%) participants wished to avoid or block as their audience (e.g., *"people who do not have similar political views."*) Abstract groups that participants wished to block from their posts' audience were highly subjective and vague, such as "*people i don't know,*" "*strangers,*", "*my enemies*" or "*everyone*" and contributed to 12.5% of participants' avoid audience preferences. Family members were often selected as a category of people participants wished to avoid as their audience for Facebook posts (15.4%). This finding is consistent with prior work that Facebook users' privacy concerns were directly associated with the number of family members a user has as friends [9]. A significant portion of people also reported that they would not mind sharing the post with anyone on their network (21%). Finally, fewer participants identified situated community based groupings (6%) or a specific person (5.3%) as audiences to avoid.

Gaps between ideal and chosen audience. Next, we examined the mapping between people's ideal audience (those they wished to target and those they wished to block) and their chosen audience. First, we observed a difference in granularity between people's ideal audience and their chosen audience for Facebook posts. Although people's ideal audience groups ranged from one specific person to broad interest/trait based groups, they all mapped to the default option, "friends on Facebook" as their chosen audience. Fewer than 5% used the custom setting on Facebook to manually include or exclude individuals in their chosen audience. Thus, consistent with prior work [8, 34], we noted gaps in granularity between users' ideal and chosen audience on Facebook.

Second, current audience selection controls on Facebook allow users to manually select individuals they would like to exclude as an audience to their posts. However, we found that participants more commonly wished to exclude broad interest/trait based groups and family members. Based on the specific constraint, the excluded audience could number anywhere between tens (e.g., "*family members*") to hundreds of people (e.g., "*friends who hate metal music.*"). In other words, excluding people from one's audience does not scale with the existing audience controls on Facebook. Similarly, manually created lists do not scale to instances where users wished to selectively target as well as block individuals as audience to their posts.

Finally, the categories of people's ideal audience might evolve and their distributions might change over time. For instance, in 2016, Litt and Hargittai [35] found that a significant portion of people's imagined audience included their personal ties (such as friends and family.) In contrast, our study shows a majority of users identifying interest/trait-based groups as their ideal audience. Furthermore, the grouping of individuals belonging to these audience categories might change with the social network of the user — the categories are dynamic and not static predictions of

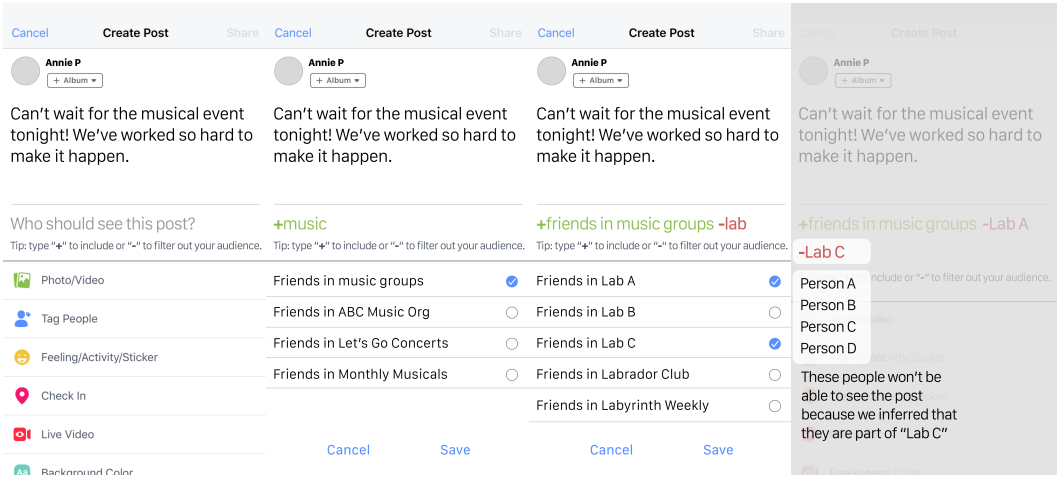


Fig. 2. A screenshot of the hi-fidelity, sensitizing DAS prototype. (a) Underneath where a user would type out the post, we added a second text box where the user could specify their audience constraints. (b-c) They can use a plus symbol ("+") for targeting or a minus symbol ("-") for filtering. Based on the user's specified audience constraints, DAS algorithmically categorizes people in their social network into relevant audience categories and displays recommendations. (d) Users can click on the audience categories generated by DAS to view who among their friends were predicted to belong to a specific constraint.

individuals into groups. Thus, audience categories might evolve and people's preferences for their ideal audience might change as platforms evolve and norms around their use change, suggesting that static list construction tools alone are likely insufficient.

The mismatch in granularity between people's ideal and chosen audience, lack of scalability of current audience selection controls, and the evolving nature of users' preferences provide credence to the *potential* utility of a dynamic audience selection tool: i.e., an intelligent tool that allows users to specify finer granularity groupings of people as their audience. In subsequent sections, we use these findings to inform our prototype design and situate our evaluative findings.

4 PROTOTYPE DESIGN

Informed by the findings from RQ1, we created a sensitizing prototype of an intelligent dynamic audience selection mechanism for Facebook. Sensitizing prototypes are instantiations of futuristic design concepts meant to catalyze further innovation [47, 63]; the contribution of such design processes are not the artifacts themselves but the design knowledge synthesized in their creation and evaluation [11]. The need for sensitizing prototypes that embody prior knowledge has been acknowledged by prior HCI research to identify design opportunities for future technology concepts [47, 63]. We used this prototype as a design probe to evaluate the potential utility and intrusiveness of a dynamic audience selection mechanism on Facebook. In short, our high-level design goals were 1) to have a simple, intuitive mechanism for dynamic audience selection, 2) to support expressive audience specification via user generated descriptors, and 3) to show algorithmic inferences based on audience specifications.

Note that our core goal was to use our sensitizing prototype as a design probe to explore the utility-intrusiveness trade-off of dynamic audience selection on Facebook, not to design an immediately implementable prototype. Still, many of the affordances we envisioned in our sensitizing prototype

can be implemented now or in the near future — indeed, platforms like Facebook already enable advertisers to target users based on highly specific categories such as interests, behavior and demographics [1]. From 2014 - 2019, Facebook also had a semantic search engine, entitled "Graph Search", that would allow users to construct granular search queries over their friend networks: e.g., "Friends who Like Star Wars and Harry Potter" [2].

We started our prototype design by parallel prototyping on paper and improving the fidelity of the prototype over several iterations. We implemented static screens in Sketch, a digital design toolkit for collaborative prototyping, to simulate the Facebook mobile interface with the idea that such an audience selection mechanism could plausibly be implemented by the social networking platform. As shown in Figure 2, underneath the Facebook composer, where a user would type out the content of their post, we added a second text box to specify their audience preferences for each post, on-the-fly, at the time of posting. Several considerations went into our prototype design:

Improving expressiveness and simplicity of audience specification. How can we design a simple, fast and intuitive interface that affords easy specification of dynamic audience preferences? To improve expressiveness, our design leveraged the popular and familiar use of special symbols such as hashtags "#" and "@" on social media that have gained wide reach [43]. We designed the prototype such that users can use a plus symbol ("+") to specify their target audience and the minus symbol ("-") for filtering or excluding individuals/groups from their post's audience. We hypothesized that the +/- notation would be familiar to the end user and consistent with people's mental models for inclusion and exclusion. With this design, we moved away from default audience settings on Facebook and let the users provide their own, natural language descriptions of their ideal audience on a post-by-post basis. For example, a user can type in "`+ friends who like to dance`" and/or "`+ friends who are currently online`" and/or "`- relatives who do not share my political views`".

In contrast to the current Facebook interface that allows either the inclusion *or* exclusion of individuals into or from one's desired audience using the custom setting, our design allowed users to select and exclude groups of people in their audience simultaneously. Furthermore, to support multiple, overlapping audience categories that we identified from RQ1 findings, we designed DAS such that users could specify multiple constraints for a single post. We also considered the prominence given to audience selection in the design of DAS. The current interface for audience selection on Facebook requires anywhere between 1-6+ clicks based on users' intentions (i.e., one click for the default option and at least six clicks for custom filtering). To make audience selection faster, simpler, and more prominent, we moved the audience selection interface next to the post composer in our prototype (see Figure 2), to couple the relationship between composing a post and selecting an audience.

Constraints on audience specification. Currently, social media platforms provide default labels (such as "friends," "family," "others," "friends of friends," "followers," etc.,) for users to articulate their social connections and select their audience for content shared on their platform. However, our findings from RQ1 demonstrated that users conceptualize and describe their audience in various ways that do not align with the audience as articulated through the platform [38]. This is consistent with prior work suggesting that static, apriori lists of people poorly map onto who people want to share content with in practice [31, 45]. Our goal was to design the prototype such that it allowed users to articulate their audience in their own words per post. We draw on findings from RQ1 to identify the following dynamic constraints that could meaningfully be supported by the prototype:

Personal attribute-based: audience constraints on context-agnostic, interest or trait-based attributes of one's audience — e.g., if a post is about a new album released by a folk band, the user might want to share it only with people interested in folk music.

Content-based: audience constraints based on the content being shared — e.g., for personal updates that might be inappropriate for a wider audience, a user might want to selectively disclose to a narrow audience.

Context-based: audience constraints based on situational factors — e.g., a user might want to share relevant information about an event for people at that event while the event was happening.

Prior work suggests that personal attribute-based and content-based audience constraints should be technically feasible by leveraging topic modeling, tie strength information between individuals, similarities between dyads' sharing behavior, and group membership [33, 66]. On the other hand, incorporating context-based situational factors may still be challenging, despite advancements in AI techniques to capture context (e.g., scene context from visual data [51]).

We designed the prototype such that the three categories of constraints were not mutually exclusive. Prior work acknowledges multiplicity [37] — i.e., that users conceive of multiple audiences through a single account, sometimes conscious of the overlap among such audiences, and suggests designs such as Venn diagrams to address the ambiguity between overlapping audiences [18]. We designed our prototype to support multiplicity such that inter-related, overlapping descriptions could be combined using one or several of the above constraints (e.g., people who live near me and like the show Sherlock Holmes.) Findings from RQ1 also revealed that users' ideal audience included those outside of their friends on Facebook (e.g. "Netflix users"). To support this use-case, we designed the prototype such that it could leverage the "Public" privacy settings on Facebook to reach people outside of one's direct friends.

Implications of automated inferences. Our prototype is designed to be the front-end of an intelligent, personalized audience selection mechanism on Facebook. Once users specify their audience preferences using the +/- notation and constraints in their own words, we envisioned a data-driven algorithm that would automatically identify people in the user's network who would belong to the specified audience category constraints and make recommendations (see Figure 2).

Current audience selection tools are static and thus deterministic. Therefore, the user would know their post's audience settings (e.g. only me, friends, friends of friends, etc.) even if they do not know who might end up seeing their post [8]. In contrast, our prototype relies on a data-driven algorithm that would allow for probabilistic audience specification, necessitating consideration of false positives and false negatives. While AI-powered methods for targeting a desired audience given user-specific constraints exist (e.g., targeted advertisements, Facebook graph search), as with any data-driven algorithms (see section 2.2.1), we expect DAS to have prediction errors. Should we prioritize minimizing false negatives (someone who should be in the audience is incorrectly excluded) or false positives (someone who should not be in the audience is incorrectly included)? Given that the prototype introduces a fairly new concept of data-driven intelligent audience recommendations, we emphasized the risks of false positives in our design. If a person was predicted to belong to both (+) and (-) categories specified by the user, our design would filter that person out of the post's audience. However, outside of such audience overlap, false positives could still exist as a result of incorrect predictions.

Another consideration that went into our design was related to errors made by the data-driven algorithm in predicting people into specified audience groups. What recourse do we provide when the algorithm makes an error? Errors made by DAS do not only affect the utility of the tool — by showing participants exactly who is predicted to belong to these dynamic categories, it introduces a channel for people to learn things about their friends they may not have otherwise known and that their friends may not want them to know. Thus, algorithmic errors might also be perceived as intrusive. To evaluate this trade-off between the utility and intrusiveness of DAS, our approach was to show the user who was predicted to belong to their specified audience groups by the algorithm (see Figure 2(d)). We considered whether the design should provide explanations for the inferred

Table 1. Summary of participants' demographics for RQ2 user study

ID	Gender	Age	Education	ID	Gender	Age	Education
1	male	18-24	bachelor's	12	female	18-24	bachelor's
2	female	18-24	bachelor's	13	female	25-34	some high school
3	male	18-24	bachelor's	14	female	18-24	some college
4	male	25-34	master's	15	male	35-44	high school grad
5	female	25-34	master's	16	male	18-24	bachelor's
6	female	18-24	bachelor's	17	female	35-44	bachelor's
7	female	18-24	bachelor's	18	female	18-24	bachelor's
8	female	18-24	bachelor's	19	male	35-44	bachelor's
9	female	18-24	bachelor's	20	female	25-34	master's
10	male	18-24	bachelor's	21	female	18-24	bachelor's
11	non-binary	18-24	bachelor's	22	male	18-24	master's
				23	male	25-34	master's

audience categories. For simplicity, in the case of incorrect predictions, we designed DAS such that users could manually fix the errors by iterating and filtering the algorithm's recommendations.

5 RQ2. EVALUATING THE UTILITY AND INTRUSIVENESS OF DAS

Using the sensitizing prototype (Section 4) as a design probe, in this section, we design a user study to evaluate the perceived utility, intrusiveness and trade-offs of the tool (RQ2).

5.1 Participants

We recruited 23 participants between August to September 2020 for the user study. Among our participants, 13 people identified as women, nine identified as men, and one person identified as non-binary. Participants were between the ages of 18 and 55 and were from the US and India. The call for participation was shared widely within and outside the authors' social networks, through both email and social media. The authors contacted potential participants with study information and an online consent form. All of our participants identified as active Facebook users — i.e., they posted or shared content on their Facebook timelines at least once in the three months prior to study participation. The majority of our participants had a bachelor's degree, five had a master's degree, and three had a high school or some college degree. Table 1 summarizes participants' demographics and shows the ID we use to refer to each participant.

Ethics and Compensation. Our study protocol was approved by an IRB. We compensated participants $10 per hour for their time, and on average they participated for 54 min in the user study.

5.2 Procedure

RQ2 inquires about the utility, intrusiveness, and the trade-offs between the two for dynamic audience selection on Facebook. To address this research question, we conducted task-based testing that involved participants using our sensitizing prototype of DAS for hypothetical audience selection scenarios on Facebook. In contrast to solution-first design, scenario-based design helps us describe how people will use DAS, rather than functional specification of the system (see Table 2 in [42]). Similar techniques have been previously applied to understand privacy nudges [58] and social media posting decisions [30]. We employed a within-subjects design in which each participant interacted with the prototype under 3 different posting scenarios. Participants took part in one

60-minute remote session. Prior to our final evaluation, we conducted a preliminary pilot study with 12 participants in Spring 2019 to test an early version of our prototype. Based on feedback from the pilot, we modified our prototype design to highlight algorithmic inferences by adding a new frame in which the names of the individuals who would be included in the dynamic audience can be seen, see 2(d). Initially, participants were presented details about the study and asked to provide informed consent. Then, they completed a demographic questionnaire and proceeded to walk through the prototype. Upon receiving participants' consent, the sessions were audio recorded and then transcribed for analysis.

5.2.1 Prototype walkthrough: Participants were provided with a walkthrough of our dynamic audience selection prototype using an example posting scenario. Due to the constraints of a remote session, participants were given access to a presentation with multiple snap shots of the prototype for a given scenario. The study facilitators then shared the description of the posting scenario and provided a walkthrough of how one might use the DAS prototype and the existing Facebook interface to post. We shared with participants how they could use our prototype for specifying an audience in their own words with the +/- notation. We also clarified that the tool would make recommendations based on the specified audience descriptions in a data-driven, intelligent manner. For comparison, they were then shown the same example posting scenario using the existing Facebook interface. Participants explored the prototype for a few minutes and we answered any questions they had about the working and design of the prototype.

5.2.2 Using DAS in hypothetical posting scenarios: The main portion of the session involved task-based testing using the prototype as a design probe and aid. We prepared hypothetical scenarios for social media posting decisions to elicit the trade-offs between utility and intrusiveness with DAS. In coming up with the posting scenarios, we drew on themes from prior work on privacy and tensions in posting on social media. For instance, for one scenario, we used context collapse [37] as the theme wherein the hypothetical user did not want multiple audiences such as friends and family to overlap into one. Similarly, we drew upon literature of the benefits and challenges of making sensitive self-disclosure on social media [20] for another scenario to identify how participants made trade-offs between utility and intrusiveness. We note that our goal was not to generate an exhaustive set of dimensions for tensions in social media posting behaviors. Instead, we selected a few representative scenarios in which audience selection is challenging to examine the utility vs. intrusiveness trade-offs. We generated three scenarios that varied along the following dimensions:

1. *Audience constraints:* Drawing on our findings from RQ1, we designed three scenarios: one scenario each for personal attributes, content-based, and context-based audience constraints. Our goal was to inquire if participants' perceptions of utility and intrusiveness varied based on these constraints. 2. *Intrusiveness*: To elicit participants' feedback on the potential intrusiveness of the algorithmic inferences that DAS may make, our three scenarios varied in the level of specificity necessary to articulate one's ideal audience. We hypothesized that extremely specific audience recommendations made by a data-driven algorithm would be perceived as more intrusive than those with less specificity.

An example for personal attribute constraints with low specificity of audience is listed below. For content-based constraints on audience selection, we took inspiration from literature on self-disclosure and context collapse to describe a protagonist who wished to share mental health resources to a targeted graduate student population but not their family members. Similarly, for a context-based scenario, we introduced multiple constraints such as the protagonist's aim for fundraising for a politically charged topic without offending people with divergent views. For a complete list of scenario descriptions, please see the supplementary materials.

> Example scenario 1. Jenny is an undergraduate student who has recently gotten into the show Criminal Minds. She's enjoyed watching the show by herself, but thinks it would be more fun to binge with friends and have discussions about the show with them. She wants to make a post to see which of her friends would be interested in doing a watchalong with her.

For each scenario, participants were asked to play the role of the protagonist. First, they were asked how they would use the current Facebook interface for the hypothetical posting scenario. Then, they were asked how they would use the prototype we designed for the scenario. Specifically, they were prompted to think about audience descriptors they would use with the +/- notation with the prototype. Then, to elicit participants' perceptions on the trade-offs between utility and intrusiveness, we shared with them three factors to consider for a tool like DAS to function as described. Specifically, we highlighted that the algorithm making audience recommendations would necessitate these considerations: 1) categorizing people in the hypothetical user's social network into the specified audience categories using data generated by social interactions on the platform; 2) processing the post's content (e.g., topic of the post) in real time, prior to posting, to recommend audience categories for content-based constraints, and 3) errors which might lead to people being wrongly included or excluded in the post's audience. For each consideration, we gave examples specific to the scenario.

Given these considerations in designing something like our prototype, we then asked participants to pick 5 words from the product desirability scale [7] that best expressed their reaction to our prototype for the specific posting scenario. Participants worked through a list of 35 adjectives and described their reactions to the considerations involved in DAS's design. We used these adjectives to probe deeper into participants' perceptions of DAS's utility and intrusiveness with followup questions. Finally, we asked participants if they preferred the current Facebook interface or DAS for the hypothetical scenario, discussing the trade-offs that made them decide one way or another.

5.2.3 Activity involving participants' posting behaviors on Facebook: Participants also took part in an activity involving their own Facebook use. The goal of this step was to let the participants engage with the prototype and analyze whether they would find it useful (or intrusive) in the context of their own posting behaviors on Facebook. For this activity, participants were asked to select an old photo on their Facebook profile. At the beginning of the session, they were asked to describe the context in which they posted that photo and the people with whom they wanted to share the photo on Facebook. By asking them to pick an old photo, we aimed to eliminate any recency bias allowing participants to reflect on the context in which they originally uploaded the photo. After noting their intended audience for the chosen photo, we returned back to this activity at the end of the overall session and asked how they would use our DAS prototype to choose their ideal audience for the photo. We also asked whether they would have preferred to use our prototype, if it existed at the time of posting.

5.2.4 Semi-structured Exit Interview. Finally, we finished the study with a brief, semi-structured exit interview where we asked participants questions about their overall impression of the prototype.

5.3 Data Analysis

Qualitative data and facilitator observations informed the major findings from this study. We used a mixed inductive and deductive approach to thematic analysis [22] in analyzing and coding the user study session interviews. Three researchers reviewed the session transcripts individually first, then collaboratively developed the final codebook. We also used other collected data (e.g., the adjectives participants selected) to provide additional insights into the experience of DAS.

5.4 Results

In this section, we describe findings from the user study as they relate to the utility, intrusiveness and trade-offs of DAS.

5.4.1 RQ 2.1 Utility of DAS.. Participants' impressions of the utility of DAS varied across posting scenarios. For the attribute-based constraints scenario (i.e., in which an undergraduate was targeting friends interested in a TV show as her audience), many participants felt that DAS could help them meet new friends or learn new things about existing friends, scaling their audience up. For more sensitive or polarizing topics (i.e., the content and context-based hypothetical scenarios), participants mentioned that DAS could empower them to speak more openly about personal topics (e.g., mental health and political ideology), and encourage others to do the same. Several participants also saw DAS as a quick way to display desired audiences that they could then manually edit, helping them save time. Across all three scenarios, participants agreed that DAS helped them reflect on their own posting behaviors and on whether they post or withhold content on Facebook currently.

In general, we found four factors that made participants feel positive about DAS's utility — enlightenment, empowerment, efficiency, reduction of context-collapse — and one factor that made participants feel negative about its utility — distrust in algorithmic inferences.

Enlightenment. Several participants noted that DAS could help them learn about new people and uncover friends who might have similar shared interests. This feeling was most related to attribute-based constraints (in which the protagonists' ideal audience was friends interested in a TV show watchalong), where DAS would help users target specific interest groups in a public conversation:

> "*It's nice to be able to connect with people that you may not know... that share the same interest in TV shows that she [the protagonist] would be into. I think it's a useful tool that she can use to meet new people and make new friends…She would be able to connect with more people.* " [P19]

For the same scenario with attribute-based constraints, a few participants were also excited by DAS's potential to introduce them to new groups or audiences, or help them learn new things about people they already knew.

> "*It's definitely valuable in that sense that it can scale beyond my understanding, especially if I've got a network of like thousand plus friends, then it can be more valuable.*" [P23]

In scenarios where the intended audience was not straightforward or the post was potentially beneficial to a broader set of people, participants found DAS to be particularly valuable due to the automated audience recommendations. P7 highlighted how the auto-generated recommendations might increase the impact of a post by reaching a wider set of audience.

> *" I think someone like Eric [the protagonist] could use the recommended groups list very easily...[in a way] he may not have thought of, because he did mention the grad students in this scenario, but there could also be other recommended groups related to mental health that he might not know about. He might also want to share with [other groups] to expand the post's impact. [P7]"*

Empowerment. Participants noted that DAS's expressiveness (i.e, specifying audience categories in their own words) and customizability (i.e., configuring audience controls) helped them feel empowered to post more and improve the engagement they received on the platform. P5 says, " *it is really empowering to, you know, be able to post on my Facebook wall like this, and being able to share something I wouldn't be doing normally"* P6 highlighted that DAS was particularly valuable for higher stakes posting scenarios as users are less inhibited to share content they otherwise would

not have. She said, *"If [DAS] wasn't an option, he [Eric, the protagonist] might be afraid to make this post, and then someone who needs to see it wouldn't be able to, so I feel like it's enabling this person to post such content."* P20 echoed similar feelings in the case of voicing controversial opinions on Facebook.

> "*I think a lot about, like, how my relatives have so different perspectives on like, Black Lives Matter, and about mental health… like they have so different values from mine and then I get really conscious about it. Like, say when I post about protests, I know how they're gonna think about it. And I sometimes even want to call them out… but I never do. If I post something more controversial, rather than just sharing news, but more like actually sharing my own opinion… I'll definitely customize [the audience].*"[P20]

P17 viewed DAS as a mechanism to improve engagement to her content on Facebook and subsequently the interactions she has on the platform. P17 said that for personal posts, she would use DAS to only include people who she believed would support her and engage positively. She believed that a tool like DAS would help curb the risk of negative feedback on her posts, as well as guarantee a positive response when she wanted support from friends. She explained,

> "*Say I'm getting married tomorrow. Now in that case…I will not [include] all my friends, all my family in that post, because you got some people on Facebook who just send you friend requests just to be nosy, so they don't care about you. They're not gonna like your posts, and they're not gonna share them, and some of them are not going to congratulate you. But certainly…the ones that have your back…yeah, you can include them. In that case, I'll know who I want in my corner.*" [P17]

Efficiency. Participants frequently mentioned that DAS would help them save time in making decisions about posting on Facebook because of its scalability (i.e., targeting large audience sizes with meaningful descriptions of constraints.) P17 shared that when she wanted to make a post celebrating all of her friends who were fathers for Father's Day, she tried to individually tag each father to make sure the post reached all of them: "*I could not tag all 500, 800 men on my Facebook…It was so time-consuming. I just tagged 25 and then…If I had the prototype, all I would have had to do was type 'fathers'.*" She believed the scalability provided by DAS's audience specification would save time in posting decisions. P3 echoed similar feelings comparing DAS to the current Facebook interface and the significant amount of time he spent on posting decisions.

> *" if you have more friends, and the topic is not quite as urgent, you would most likely use your prototype, just because it greatly simplifies and speeds up the process of making a post, it's no longer taking 20 minutes to dig through your friends to figure out who you can share it with." [P3]*

A few participants also saw DAS more as a way to search for groups that they could filter more easily themselves. For example, P16, who chose not to post certain content on Facebook because "*it's not worth the time to even filter those people out.*", shared:

> "*With the prototype, I can just add people like an overarching grad student thing and get a lot of people in right away. Then I can just filter out the people I don't want in that specific list. With the old Facebook way, I would still have to filter all the people, so doing it like this [with DAS] is just a shorter list to filter.*" [P16]

Reduction of context collapse. When participants were asked about the context of previous posts or how they might use our prototype to make posts, several people answered that they already use Facebook in a very intentional or specific way. P11 outlined how every post they made was carefully thought out, filtering out posts that might not be relevant to everyone on their friends' list: "*I will only really post something on Facebook if, like, I've previously thought about it, and it's*

something that I want to, like, tell everybody who I'm Facebook friends about" (P11). Others explained how multiple, overlapping social circles hindered them from fully expressing themselves on the platform, and easily removing these contacts from a post might encourage more activity.

> "*Sometimes I intentionally won't like things or won't be posting... because I know it will share to my feed. And there are things that I enjoy but I just would rather not have my elementary school third grade teacher see it. So, I feel like... it could be really helpful... if I had the prototype [I might] post more on Facebook*" [P8]

A few participants felt that their fear of being repetitive or oversharing on Facebook would be mitigated by dynamic audience selection. P6, who had distinct social circles on Facebook – family members and friends around their age, felt that having DAS would result in more posting. They said,

> "*I do think that this prototype would, like, now make me post more things, because I kind of avoid just posting things like for club events and advertising events and like club stuff because, again, most of my friends on Facebook are family members and so they don't need to see that stuff. So, if I could exclude them from those posts, then I would definitely post more.*" [P6]

While our goal with DAS was not to facilitate increased posting behaviors on Facebook, we find that it helped participants feel more confident and comfortable about the audience of the content they might post on Facebook.

Distrust in algorithmic inferences.

When it came to deciding between Facebook and DAS in the context of the specific posting scenarios we presented, some participants expressed that the unpredictability and inconsistency of DAS outweighed its benefits. Specifically, when presented with the fact that an algorithm would be sorting people into different categories and could make errors in doing so, many participants displayed a lack of trust in the predictions the algorithm could make. The possibility of error made many participants double check DAS's recommendations to see who was in each audience group: "*You can't trust [the algorithm] completely; you have to take the time to make sure it's right as well*" (P6). Knowing that a user's interests might not be directly reflected in the kind of content they consume or share on a platform further fueled this doubt.

> "*Sometimes people might like things, because they know someone who's part of it rather than actually like being interested in whatever the post was... So I think that might make it a little bit inconsistent. [With DAS] I think you need to at least go through the list yourself and make sure that like these are the people you want to target*" [P6]

Participants were also hesitant to trust an algorithm for more nebulous categories, like political affiliation and stances on polarizing issues. While some participants noted that algorithm might have access to more data than they might, they questioned whether an algorithm could accurately pinpoint someone's opinion on various political issues, since these beliefs may be prone to changing over time.

> "*Something like a political association is not flexible, but constantly moving, a specific algorithm is kind of just placing people into categories. So it's there's not much wiggle room there... [the] political spectrum is huge,*" (P2)

Algorithmic errors also diminished the positive aspects of DAS. For instance, consider P21 who shared that DAS would be time-saving for the scenario in which the protagonist wished to share about mental health with graduate students, but not with professors and family members. When asked if she would pick the current Facebook interface or DAS for this scenario, P21 said she would not select DAS because of the manual labor involved in manually checking the inferred audience

groups by DAS. In case of algorithmic errors (e.g., a family member is predicted as a grad student), she would have to fix DAS's errors before posting. In other words, fixing errors might end-up being more time-consuming than time-saving.

> *" [I'll pick] probably Facebook as it is today. it's like a close neck and neck, but I think just the only thing that might send me towards Facebook is again just that idea of, like, kind of the overlap [of +/- audience constraints with DAS] and seeming like perhaps that that's more kind of time consuming or it could be to check everyone versus just like if it was a smaller group, I could just manually manually add them out, or restrict.[P21]"*

Similarly, consider the expressiveness and customizability of DAS enabled by algorithmic prediction of audience groups. Many participants found these features useful and empowering. However, the possibility of errors and distrust in the algorithm's predictions made participants feel that DAS was inconsistent and unpredictable. P5 described this saying, *"it is very helpful that the prototype auto-generates the categories. However, I am not sure if it will consistently predict the same 10 people when I type in '+family'. you know, with Facebook custom setting, the list is static, so I know it will always stay the same."* In other words, the potential unpredictability of DAS diminished its potential utility.

Some participants preferred posting to closed Facebook groups instead of posting on their own timeline and using DAS to find audiences, since the memberships of these groups are selective and are not burdened by DAS's uncertainty:

> *"I would just use groups that are local and politically-leaning toward whatever [protagonist] wants… I would post there. I would say those groups are an equal, if not better, smaller filtered list [compared to DAS], because the people in that group probably joined of their own volition… They're like, kind of pre-vetted because the people that are in it have to join on their own or are invited and can leave any time." [P16]*

P4 echoed similar feelings: "*With predefined lists, I know where this post is going [...] I don't know if the algo would be consistent like a group.*" For posts with more sensitive content, some utility of DAS was lost, as algorithmic doubt caused participants to double check who was placed into each category.

> *"to put complete trust in that algorithm is, uh, a little bit difficult even if you do have the option to go through and double check everything, which in it of itself would like take away from the benefit of being fast.*" [P11]

5.4.2 RQ 2.2 Intrusiveness of DAS.

Privacy-intrusive side channels. While some participants were excited about DAS's potential for introducing them to new friends and helping them learn more about existing ones, others were concerned that it could reveal things about their friends that they did not want to know, or reveal things about themselves that they did not consent to sharing. For example, P16 noted that it would be creepy if DAS could correctly infer their ideal audience if they queried for "*girl that I have a crush on*". Likewise, P4 mentioned that he wouldn't want an algorithm automatically sorting his social circle into personal categories. P4 also expressed concern about DAS inadvertently revealing personal information about his friends that he would rather hear from his friends directly. This concern made by P4 is the introduction of new side channels through which the algorithm could reveal sensitive, private information about their friends and themselves.

> "*I don't want an algorithm to label my family, I want myself to label my family[...]someone might have a nuanced opinion about [a polarizing topic]. And an algorithm making judgement calls and labeling people based on ideology seems wrong to me.*" [P4]

Worth noting, however, is that not all participants who acknowledged these side-channels felt negatively about the intrusiveness of DAS. For example, P20 mentioned that she would be shocked but curious if she found out that one of her friends had a political viewpoint that was very different from her own. P20 described how she would be interested in finding new information about her friends and how an algorithm made certain predictions about her friends – essentially appropriating DAS's inferences to learn new information that would otherwise be unavailable to her.

> "*I think I'd actually go to their profile and check, like, 'What do they post that they're categorized this way?' If an algorithm told me that my friend [had controversial views] and it was way later into knowing them, I would be pretty shocked.*" [P20]

Other participants expressed apathy towards the intrusiveness of DAS. For instance, P10 noted that being sorted into categories based on his content was a little strange, but he would be okay with it overall.

> "*[DAS] does give me a little hesitation because it's like do I want to be grouped as liberal just because I like left-winged posts .... definitely like politics I think is where it gets me the most. But overall I don't think it would really bother me....I think it's useful, but that's also because I am like, I'm on the end where like, I don't really care about my data as much as long as you can't see my credit card number. I don't really care what you tell me. So yeah, that's my thing.*" [P10]

Similarly, P11 shared that the reason they did not feel strongly about intrusiveness was due to the prevalence of data collection and usage for ads by social networking platforms. P11 said, *"I guess a little bit like viscerally weird to like see it, but I mean, reasonably, I know that all of that data is generally already being collected. Anyways. So it doesn't matter that much to me."*

One participant noted that the inherent reflexivity and iterative nature of DAS could allow people to reverse engineer what data on their friends' profiles made DAS predict them as belonging to a certain audience category. P21 says "*Wow, [...] if you think something like [people's political ideology] might be so deeply correlated with [their] profile factors and can be predicted [..] that's kind of intimidating. it goes back to how [is DAS] getting this inference about my friends? If there's, to me, no information that would give you that categorization.*"

5.4.3 RQ 2.3 Trade-offs between utility and intrusiveness of DAS.

We found that the stakes of a posting decision defined the trade-offs between the utility and intrusiveness of DAS. Participants described the stakes in terms of sensitivity of the content, potential risks of unintended audience seeing the post, and potential benefits of the post reaching its intended audience. Several participants made a decision between using DAS and Facebook's existing audience controls based on the stakes involved. We found that when posting about more personal, sensitive topics (mental health, politics, etc.), participants were more likely to question the accuracy of the algorithm and, in turn, were more hesitant to use DAS. One participant alluded to the stressful impact of a wrongful prediction.

> "*if it does mess up something, a post that is supposed to be private—It's like a really big problem, right? And that's, and that's very scary to think about. If it's a private post, I don't want the world to see.*" [P10]

Faced with using DAS or Facebook's current interface for high stakes posting decision, many participants opted for Facebook's current interface, preferring to manually filter out unwanted audiences instead of relying on an error-prone algorithm.

> "*So, if you were incredibly serious and it's something that's like really personal related to life or death, you would probably use Facebook's current interface to guarantee that the algorithm didn't miss anything.*" [P3]

Conversely, while the probabilistic nature of DAS made some participants hesitant to use it for high-stakes scenarios, others found it difficult to justify the potential for privacy intrusiveness entailed by DAS for low stakes scenarios — i.e., scenarios where errors in audience selection would not be a significant concern. Some users did not feel the need to target or exclude audiences from posts, so having the algorithm scan and suggest audiences felt intrusive and unnecessary. P3, when asked how they would react to DAS if it had been available when they posted some pictures from a vacation, stated,

> "*Because this photo is not something that's sensitive...I think having an interface or having your prototype, analyze, you know everything about a photo and kind of who interacts would be a little intrusive because I don't gain that much out of your prototype.*" [P3]

Beyond the stakes of a given post, participants also considered the platform providing the service in weighing the potential utility versus intrusiveness of DAS. For example, P23 expressed that despite the potential utility that DAS provides, it would be a "hard sell" for platforms like Facebook to justify the intrusiveness of a tool like DAS given that they might use those same algorithms for secondary uses like ad targeting.

> "*So, if Facebook can, as a platform can show that you know what you're trying to understand these [audience categories] in order to protect you. Not in order to sell you things, then I think it's okay. But that's a hard sell. [...] I'm okay with that. [...] I would much rather have better ads [...] but [other] people [might not be] okay with it." [P23]*

In summary, we found that for low-stakes posting decisions, some participants appreciated DAS's utility, but found it difficult to justify the intrusiveness involved in algorithmic predictions. On the other hand, for sensitive, high stakes posting decisions, while participants felt that the customizability offered by DAS was empowering, the introduction of privacy intrusive side channels discouraged them from using DAS. Moreover, participants believed that the algorithms powering DAS would likely also be used for other additional purposes, like ad targeting — in this case, intrusiveness also outweighed utility.

6 DISCUSSION

Our findings suggest that many participants found utility in using DAS. Its expressiveness, customizability, and scalability helped people feel more comfortable sharing content that they wanted to share, empowered them to improve engagement with their posts, and saved time in making posting decisions. These findings strengthen findings from prior work suggesting that more granular audience controls would help people self-censor less and reduce interpersonal privacy violations due to context collapse and inappropriate access. However, distrust in the algorithms powering DAS and the uncertainty associated with prediction errors reduced its utility. Moreover, owing to the privacy-intrusive side channels introduced by DAS, many participants found the concept to be intrusive. On balance, participants perceptions' of low reliability and intrusiveness outweighed the utility they saw in DAS, for both low-stakes and high-stakes posting scenarios.

Drawing on these findings, we next reflect on whether DAS is an appropriate design solution for audience selection on Facebook. Baumer and Silberman [6] present three questions to articulate when new technology may be inappropriate. While the goal of these questions is to identify when not to design, we adopt these questions to reflect on whether deploying DAS for Facebook is an appropriate design solution and how lessons from this study can inform future designs.

6.1 Is there an equally viable low-tech or no-tech approach to the situation?

Our DAS prototype envisioned an intelligent, personalized, on-the-fly tool through which users can specify constraints for a Facebook audience in natural language, on a post-by-post basis. We

found that participants found utility in our prototype, owing to its expressiveness, customizability and scalability. However, participants also found DAS to be intrusive.

A no-tech intervention to audience selection on Facebook would involve using current, static audience control mechanisms for custom settings and manually curated lists. However, prior work suggests that these controls poorly map onto users' *imagined* audiences [8], and our first study suggests that these controls rarely map onto users' *ideal* audiences. Thus, it appears what's needed may be something in-between purely static audience selection controls and the fully customizable DAS solution we evaluated in this work — i.e., a low-tech, hybrid approach.

Applying DAS to low-stakes vs. high stakes posting scenarios. The three hypothetical scenarios we designed for RQ2 varied in the audience constraints and specificity necessary to express one's ideal audience. Our evaluation, however, revealed an additional factor along which the three scenarios were different: the stakes of misspecifying one's ideal audience. Sometimes, participants questioned whether an intrusive tool like DAS was justified for a low stakes scenario, like posting about a television show watch-along. A significant majority, however, found DAS to be useful for low stakes posting scenarios. In contrast, for high stakes scenarios (e.g., posting sensitive or polarizing content), participants wanted DAS to be reliable and predictable. Thus, the potential for algorithmic errors lowered DAS's utility across these scenarios.

If DAS, as we conceived of it, is too intrusive for low-stakes scenarios, but too unpredictable for high-stakes scenarios, when might DAS be appropriate — i.e., when might its utility outweigh its intrusiveness?

A low-tech approach here might alleviate concerns about intrusiveness while also providing reliability and consistency in audience groups. For example, one possible "low-tech" design could use intelligent sorting algorithms to auto-generate pre-defined lists of audience groups based on user-generated constraints, instead of predicting audience groups on the fly per post. This approach presents a middle ground between current static lists and DAS's dynamic audience recommendations, potentially reducing the intrusiveness experienced on a post-by-post basis. Instead, the curated list allows users to fix algorithmic errors once and affords consistency and predictability for future use. For example, we can envision adopting DAS to create a list for `'+(graduate students)'` for one of our hypothetical scenarios. Once DAS algorithmically groups people in the protagonist's network into this group, the protagonist can then save this custom grouping as part of their audience controls and use it when required. While the re-usability of such audience groups remains unknown, our participant perceptions suggest that repeated use of some audience constraints is useful. Prior work shows that it is technically feasible to to make privacy policy recommendations on social networking sites based on past sharing behaviors [32, 39]. Future work can employ similar techniques to understand the re-usability of DAS generated audience groups. Furthermore, this approach would allow users to choose when they wish to leverage the predictive power of DAS for their audience selection.

For low-stakes scenarios, we found that although participants appreciated the utility of DAS, they questioned whether the intrusiveness involved in algorithmic predictions was justified. Users might find DAS less intrusive and more useful for low-stakes posting scenarios if they were able to trust that platforms would use the output of DAS to provide better audience controls and not for secondary uses (such as targeted ads). Many social networking platforms are already using the sorts of inferences that might power DAS to serve users targeted advertisements, but a tool like DAS would bring privacy concerns related to these inferences to the forefront of users' minds.

Despite their reservations about platforms like Facebook using the predictive capabilities of DAS for secondary uses like ad targeting, participants also expressed a desire for social networking platforms to use predictive inferences to prevent anti-social behaviors — ostensibly, a secondary use. When thinking about who they would like to exclude from their ideal audience, one participant

wanted to write "*stalkers*." Upon reflection, however, they stated: "*if FB knows someone is say [a] stalker, then it should do something about it.*" This example illustrates how the reflexive nature of DAS could make end-users more conscious of the broader implications of algorithmic inferences of personal attributes — an end that may be generally beneficial for social media users, but may be oppositional to the short-term goals of the social media platforms.

The concept of stakes in the context of audience selection on social networking sites is understudied. Different from tensions or trade-offs that people navigate while posting on social media, stakes describe situations in which misspecified audiences could potentially serve to harm users. Our findings suggest that stakes play a significant moderating role for dynamic audience selection preferences on Facebook. Future work might taxonomize different posting scenarios and their stakes more deeply, and explore how dynamic audience selection might apply to different contexts.

6.2 Might deploying the technology result in more harm than the situation the technology is meant to address?

We identified two negative consequences in using DAS for audience selection on Facebook: the introduction of privacy-intrusive side channels and algorithmic errors.

Introduction of privacy-intrusive side channels. In the social networking context, we define a privacy-intrusive side-channel as a means of inferring personal information about one's connections without explicit consent through artifacts of the social networking platform. Our findings suggest that DAS may introduce such side-channels through which users can infer information about friends' interests and attributes via their interaction with DAS. Some participants were uncomfortable to learn new things about their social connections from an algorithm. That the algorithm would make inferences on one's social connections based on data to which they did not have access further complicated how participants felt about DAS. How can we prevent unintended side channels as a result of dynamic audience controls? Should audience information as described through DAS be revealed in the post? A fruitful direction for future work might be to explore means of limiting side-channel inferences while preserving the utility of a dynamic audience selection tool. One way this could be done is by limiting inferences only to information understood to be public — e.g., demographic and interest information explicitly articulated in one's profile; this limitation would, in turn, limit the utility of DAS as well.

Recourse for algorithmic errors. Algorithmic errors made some users perceive DAS as unpredictable. However, more accurate predictions of people's network into dynamic audience categories could be a double-edged sword. Some participants did not mind predictions based on others' personal information, and said accurate predictions would make DAS more reliable and useful (P23). Other participants felt that platforms making predictions about people on their network was highly intrusive, despite the utility provided by DAS (P4). At what level of accuracy does DAS cross the line from being more useful to more intrusive? What data streams would users find acceptable for DAS to use in order to make inferences about their friends? Following the exploratory analysis with our sensitizing prototype, future work could explore more concretely testable hypotheses.

6.3 Does the technology solve a computationally tractable problem rather than address an actual situation?

The underlying problem that DAS aimed to address — to bridge the mismatch between users' intended audience and their chosen audience on Facebook — is indeed real. For example, there is ample prior work illustrating privacy violations and sharing regret on social networking services owing to context collapse and boundary regulation [3, 8, 37, 45, 53, 56].

Moreover, the results from our survey suggest that there is a significant misalignment between people's ideal audience and the audience they select using static controls. Indeed, our prototype design was informed by the gaps we identified between participants' ideal and chosen audience on Facebook. Furthermore, in identifying audience categories from participants' descriptions of their ideal audience, we extended beyond simplified articulation of relationships on platforms (like "friends" or "followers").

In short, we identified a need for more nuanced articulations of audience categories. However, our evaluations suggest that systems that attempt to help users intelligently articulate their ideal audiences on social media must do so in a manner that: (i) is predictable and reliable; and, (ii) protects against the accidental divulging of personal information that users might want hidden. Moreover, since DAS is reflexive and foregrounds the power of AI in profiling individuals, social networking platforms should be prepared to be transparent in explaining how such inferences are made — advances in explainable AI may be helpful for these purposes for future iterations of DAS.

7 LIMITATIONS

The generalizability of our findings is limited by selection bias. Participants in the survey study were mostly 18-24 years old, and more than 90% of them had more than seven years of internet experience. They also came from predominantly English-speaking, western populations. Participants in the user study for RQ2 also belonged to a similar demographic. For both studies, we recruited active Facebook users; they posted at least once in the three months before their participation. Therefore, our sample is not representative of the national and international demographic, adults over 35 years old, and Facebook users who are less or more active on the platform. We also did not account for characteristics such as participants' social network and social media use that might play a role in audience selection — for instance, how often they post, the number of friends and social interactions they have on platform. Lastly, external factors beyond our observation might have affected participants' feelings about the utility and intrusiveness of DAS. We did not collect data on participants' prior experiences with privacy violations on Facebook, their boundary regulation strategies or awareness or use of audience controls on Facebook. These additional factors might impact how useful or intrusive users found DAS. Future work can explore connections between platform use, general privacy perceptions and responses to dynamic audience selection mechanisms. Despite these limitations, we present useful insights into the utility and intrusiveness of dynamic audience selection for Facebook that have implications for the future design of audience controls for social networking sites.

We designed our sensitizing prototype as one version of how dynamic audience selection might look like on Facebook. Our findings on the utility and intrusiveness of DAS, although novel, are exploratory in the context of our specific prototype. Future research on the design of audience controls for social networking sites can examine additional considerations we highlight (such as posting stakes) and design mechanisms that maintain the utility of dynamic audience selection, while lowering intrusiveness. We also focus on one social media platform, Facebook. Prior work shows that users think about affordances of a social media platform, particularly related to audience and content sharing, in the context of all of the communication tools available to them [65] and they combine multiple channels to create composite sharing features when any one platform is insufficient [46]. Therefore, a fruitful direction for future work is to explore the utility, intrusiveness and trade-offs involved in intelligent audience controls for selective sharing across an ecology of social media platforms.

8 CONCLUSION

In this paper, we designed a sensitizing prototype for dynamic audience selection on Facebook and explored the utility versus intrusiveness trade-off of such a tool. Our findings suggest that participants found utility in DAS's expressiveness, customizability, and efficiency, and felt more comfortable sharing content on Facebook with our sensitizing prototype. However, the unpredictability and inconsistency of algorithmic predictions reduced this utility. Furthermore, distrust in algorithmic predictions and the emergence of privacy-intrusive side channels made many participants find DAS to be intrusive. Our study showed that the stakes of posting decisions played an important moderating role in how participants viewed the trade-offs between utility and intrusiveness. On balance, the intrusiveness of DAS may outweigh its utility for real-time audience selection; however, there may be variations of our prototype design that strike a better balance — e.g., for helping users construct pre-defined audience lists. In short, with these findings, we extend current audience controls for social networking platforms and suggest design considerations for future intelligent audience selection tools.

ACKNOWLEDGMENTS

This research was generously supported, in part, by the National Science Foundation through grant CRII:SaTC-1755625. We thank Dr. Elissa Redmiles, students in the Georgia Tech Usable Privacy and Security class, and our anonymous reviewers for their valuable feedback on the prototype, the study design, and this paper.

REFERENCES

[1] [n.d.]. https://www.facebook.com/business/ads/ad-targeting
[2] [n.d.]. https://social.techcrunch.com/2013/01/15/facebook-announces-its-third-pillar-graph-search/
[3] Alessandro Acquisti and Ralph Gross. 2006. Imagined communities: Awareness, information sharing, and privacy on the Facebook. In *International workshop on privacy enhancing technologies*. Springer, 36–58.
[4] Davide Alberto Albertini, Barbara Carminati, and Elena Ferrari. 2016. Privacy settings recommender for online social network. In *2016 IEEE 2nd International Conference on Collaboration and Internet Computing (CIC)*. IEEE, 514–521.
[5] Saleema Amershi, James Fogarty, and Daniel Weld. 2012. Regroup: interactive machine learning for on-demand group creation in social networks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. Association for Computing Machinery, 21–30. https://doi.org/10.1145/2207676.2207680
[6] Eric PS Baumer and M Six Silberman. 2011. When the implication is not to design (technology). In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2271–2274.
[7] Joey Benedek and Trish Miner. 2002. Measuring Desirability: New methods for evaluating desirability in a usability lab setting. *Proceedings of Usability Professionals Association* 2003, 8-12 (2002), 57.
[8] Michael S Bernstein, Eytan Bakshy, Moira Burke, and Brian Karrer. 2013. Quantifying the invisible audience in social networks. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 21–30.
[9] Jens Binder, Andrew Howes, and Alistair Sutcliffe. 2009. The problem of conflicting social spheres: effects of network structure on experienced tension in social network sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. Association for Computing Machinery, 965–974. https://doi.org/10.1145/1518701.1518849
[10] Herbert Blumer. 1954. What is wrong with social theory? *American sociological review* 19, 1 (1954), 3–10.
[11] Glenn A Bowen. 2006. Grounded theory and sensitizing concepts. *International journal of qualitative methods* 5, 3 (2006), 12–23.
[12] David Russell Brake. 2012. Who do they think they're talking to? Framings of the audience by social media users. *International journal of communication* 6 (2012), 21.
[13] Eliane Bucher, Christian Fieseler, and Anne Suphan. 2013. The stress potential of social media in the workplace. *Information, Communication & Society* 16, 10 (2013), 1639–1667.
[14] Moira Burke and Robert E Kraut. 2014. Growing closer on facebook: changes in tie strength through social network site use. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 4187–4196.
[15] Gorrell P Cheek and Mohamed Shehab. 2012. Policy-by-example for online social networks. In *Proceedings of the 17th ACM symposium on Access Control Models and Technologies*. 23–32.

[16] Ramnath K Chellappa and Raymond G Sin. 2005. Personalization versus privacy: An empirical examination of the online consumer's dilemma. *Information technology and management* 6, 2-3 (2005), 181–202.

[17] Sauvik Das and Adam Kramer. 2013. Self-censorship on Facebook. In *Seventh international AAAI conference on weblogs and social media*.

[18] Serge Egelman, Andrew Oates, and Shriram Krishnamurthi. 2011. Oops, I did it again: mitigating repeated access control errors on facebook. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, 2295–2304. https://doi.org/10.1145/1978942.1979280

[19] Nicole Ellison, Rebecca Gray, Jessica Vitak, Cliff Lampe, and Andrew T Fiore. 2013. Calling all Facebook friends: Exploring requests for help on Facebook. In *Seventh international AAAI conference on Weblogs and social media*. Citeseer.

[20] Sindhu Kiranmai Ernala, Asra F Rizvi, Michael L Birnbaum, John M Kane, and Munmun De Choudhury. 2017. Linguistic markers indicating therapeutic outcomes of social media disclosures of schizophrenia. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–27.

[21] 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 News Feeds. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 153–162.

[22] Jennifer Fereday and Eimear Muir-Cochrane. 2006. Demonstrating Rigor Using Thematic Analysis: A Hybrid Approach of Inductive and Deductive Coding and Theme Development. *International Journal of Qualitative Methods* 5, 1 (Mar 2006), 80–92. https://doi.org/10.1177/160940690600500107

[23] S Frederic and H Woodrow. 2012. Boundary regulation in social media. In *Proc. CSCW*, Vol. 769.

[24] Eric Gilbert. 2012. Designing social translucence over social networks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2731–2740.

[25] Erving Goffman et al. 1978. *The presentation of self in everyday life*. Harmondsworth London.

[26] Anatoliy Gruzd, Barry Wellman, and Yuri Takhteyev. 2011. Imagining Twitter as an Imagined Community. *American Behavioral Scientist* 55, 10 (Oct 2011), 1294–1318. https://doi.org/10.1177/0002764211409378

[27] Andrew F Hayes, Dietram A Scheufele, and Michael E Huge. 2006. Nonparticipation as self-censorship: Publicly observable political activity in a polarized opinion climate. *Political Behavior* 28, 3 (2006), 259–283.

[28] Bernie Hogan. 2010. The presentation of self in the age of social media: Distinguishing performances and exhibitions online. *Bulletin of Science, Technology & Society* 30, 6 (2010), 377–386.

[29] Sanjay Kairam, Mike Brzozowski, David Huffaker, and Ed Chi. 2012. Talking in circles: selective sharing in google+. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1065–1074.

[30] Harmanpreet Kaur, Cliff Lampe, and Walter S Lasecki. 2020. Using affordances to improve AI support of social media posting decisions. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*. 556–567.

[31] Patrick Gage Kelley, Robin Brewer, Yael Mayer, Lorrie Faith Cranor, and Norman Sadeh. 2011. An investigation into facebook friend grouping. In *IFIP Conference on Human-Computer Interaction*. Springer, 216–233.

[32] A Can Kurtan and Pınar Yolum. 2021. Assisting humans in privacy management: an agent-based approach. *Autonomous Agents and Multi-Agent Systems* 35, 1 (2021), 1–33.

[33] Qingrui Li, Juan Li, Hui Wang, and Ashok Ginjala. 2011. Semantics-enhanced privacy recommendation for social networking sites. In *2011IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications*. IEEE, 226–233.

[34] Eden Litt. 2012. Knock, knock. Who's there? The imagined audience. *Journal of broadcasting & electronic media* 56, 3 (2012), 330–345.

[35] Eden Litt and Eszter Hargittai. 2016. The Imagined Audience on Social Network Sites. *Social Media + Society* 2, 1 (Jan 2016), 205630511663348. https://doi.org/10.1177/2056305116633482

[36] Naresh K. Malhotra, Sung S. Kim, and James Agarwal. 2004. Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. *Information Systems Research* 15, 4 (Dec 2004), 336–355. https://doi.org/10.1287/isre.1040.0032

[37] Alice E Marwick and Danah Boyd. 2011. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New media & society* 13, 1 (2011), 114–133.

[38] Alice E Marwick and Danah Boyd. 2014. Networked privacy: How teenagers negotiate context in social media. *New media & society* 16, 7 (2014), 1051–1067.

[39] Gaurav Misra and Jose M Such. 2017. Pacman: Personal agent for access control in social media. *IEEE Internet Computing* 21, 6 (2017), 18–26.

[40] Bernardo Reynolds, Jayant Venkatanathan, Jorge Gonçalves, and Vassilis Kostakos. 2011. Sharing ephemeral information in online social networks: privacy perceptions and behaviours. In *IFIP Conference on Human-Computer Interaction*. Springer, 204–215.

[41] M. D. Roblyer, Michelle McDaniel, Marsena Webb, James Herman, and James Vince Witty. 2010. Findings on Facebook in higher education: A comparison of college faculty and student uses and perceptions of social networking sites. *The Internet and Higher Education* 13, 3 (Jun 2010), 134–140. https://doi.org/10.1016/j.iheduc.2010.03.002

[42] Mary Beth Rosson and John M Carroll. 2009. Scenario based design. *Human-computer interaction. boca raton, FL* (2009), 145–162.

[43] Jeffrey M Rzeszotarski, Emma S Spiro, Jorge Nathan Matias, Andrés Monroy-Hernández, and Meredith Ringel Morris. 2014. Is anyone out there? Unpacking Q&A hashtags on Twitter. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2755–2758.

[44] Meredith M. Skeels and Jonathan Grudin. 2009. When social networks cross boundaries: a case study of workplace use of facebook and linkedin. In *Proceedings of the ACM 2009 international conference on Supporting group work (GROUP '09)*. Association for Computing Machinery, 95–104. https://doi.org/10.1145/1531674.1531689

[45] Manya Sleeper, Rebecca Balebako, Sauvik Das, Amber Lynn McConahy, Jason Wiese, and Lorrie Faith Cranor. 2013. The post that wasn't: exploring self-censorship on facebook. In *Proceedings of the 2013 conference on Computer supported cooperative work*. 793–802.

[46] Manya Sleeper, William Melicher, Hana Habib, Lujo Bauer, Lorrie Faith Cranor, and Michelle L Mazurek. 2016. Sharing personal content online: Exploring channel choice and multi-channel behaviors. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 101–112.

[47] Erik Stolterman and Mikael Wiberg. 2010. Concept-driven interaction design research. *Human–Computer Interaction* 25, 2 (2010), 95–118.

[48] Fred Stutzman and Jacob Kramer-Duffield. 2010. Friends only: examining a privacy-enhancing behavior in facebook. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. Association for Computing Machinery, 1553–1562. https://doi.org/10.1145/1753326.1753559

[49] Kaveri Subrahmanyam, Stephanie M. Reich, Natalia Waechter, and Guadalupe Espinoza. 2008. Online and offline social networks: Use of social networking sites by emerging adults. *Journal of Applied Developmental Psychology* 29, 6 (Nov 2008), 420–433. https://doi.org/10.1016/j.appdev.2008.07.003

[50] Omer Tene and Jules Polonetsky. 2013. A theory of creepy: technology, privacy and shifting social norms. *Yale JL & Tech.* 16 (2013), 59.

[51] Ashwini Tonge, Cornelia Caragea, and Anna Squicciarini. 2018. Uncovering scene context for predicting privacy of online shared images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32.

[52] Amin Tootoonchian, Stefan Saroiu, Yashar Ganjali, and Alec Wolman. 2009. Lockr: better privacy for social networks. In *Proceedings of the 5th international conference on Emerging networking experiments and technologies (CoNEXT '09)*. Association for Computing Machinery, 169–180. https://doi.org/10.1145/1658939.1658959

[53] Zeynep Tufekci. 2008. Can you see me now? Audience and disclosure regulation in online social network sites. *Bulletin of Science, Technology & Society* 28, 1 (2008), 20–36.

[54] Blase Ur, Pedro Giovanni Leon, Lorrie Faith Cranor, Richard Shay, and Yang Wang. 2012. Smart, useful, scary, creepy: perceptions of online behavioral advertising. In *proceedings of the eighth symposium on usable privacy and security*. 1–15.

[55] Blase Ur, Pedro Giovanni Leon, Lorrie Faith Cranor, Richard Shay, and Yang Wang. 2012. Smart, useful, scary, creepy: perceptions of online behavioral advertising. In *Proceedings of the Eighth Symposium on Usable Privacy and Security (SOUPS '12)*. Association for Computing Machinery, 1–15. https://doi.org/10.1145/2335356.2335362

[56] Jessica Vitak. 2012. The impact of context collapse and privacy on social network site disclosures. *Journal of broadcasting & electronic media* 56, 4 (2012), 451–470.

[57] Jessica Vitak, Stacy Blasiola, Sameer Patil, and Eden Litt. 2015. Balancing Audience and Privacy Tensions on Social Network Sites: Strategies of Highly Engaged Users. *International Journal of Communication* 9, 0 (May 2015), 20.

[58] Yang Wang, Pedro Giovanni Leon, Alessandro Acquisti, Lorrie Faith Cranor, Alain Forget, and Norman Sadeh. 2014. A field trial of privacy nudges for facebook. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2367–2376.

[59] Jason Watson, Michael Whitney, and Heather Richter Lipford. 2009. Configuring audience-oriented privacy policies. In *Proceedings of the 2nd ACM workshop on Assurable and usable security configuration (SafeConfig '09)*. Association for Computing Machinery, 71–78. https://doi.org/10.1145/1655062.1655076

[60] Pamela Wisniewski, Heather Lipford, and David Wilson. 2012. Fighting for my space: Coping mechanisms for SNS boundary regulation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 609–618.

[61] Sijia Xiao, Danaë Metaxa, Joon Sung Park, Karrie Karahalios, and Niloufar Salehi. 2020. Random, Messy, Funny, Raw: Finstas as Intimate Reconfigurations of Social Media. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[62] Heng Xu, Xin Robert Luo, John M Carroll, and Mary Beth Rosson. 2011. The personalization privacy paradox: An exploratory study of decision making process for location-aware marketing. *Decision support systems* 51, 1 (2011),

42–52.

[63] Qian Yang. 2018. Machine learning as a UX design material: How can we imagine beyond automation, recommenders, and reminders?. In *AAAI Spring Symposia*.

[64] Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. 2020. Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, 1–13. https://doi.org/10.1145/3313831.3376301

[65] Xuan Zhao, Cliff Lampe, and Nicole B Ellison. 2016. The social media ecology: User perceptions, strategies and challenges. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 89–100.

[66] Haoti Zhong, Anna Cinzia Squicciarini, David J Miller, and Cornelia Caragea. 2017. A Group-Based Personalized Model for Image Privacy Classification and Labeling.. In *IJCAI*, Vol. 17. 3952–3958.

A APPENDIX

A.1

Table 2. Survey questionnaire for RQ1

Question text	Response format	Response choices
Block 0		
1. Do you currently own a Facebook account?	multiple choice	• Yes, I own a Facebook account • No, I do not own a Facebook account
Block 1: Learning about participants' most recent post		
2. Please look back on your Facebook timeline. When was the last time you posted a status update or shared someone else's posts?	date	
3. How would you describe the nature of the content? Please select all that apply.	multiple choice	• Entertainment (*ex: 21 Savage is lit*) • Politics (*ex: Please rethink student #loanforgiveness for the 2020 elections!*) • Personal updates (*ex: excited to share that I'll be starting my PhD!*) • Personal opinion (*ex: pineapples shouldn't be on pizzas)* • Other (please specify)
4. Think about the context in which you published the post. In your own words, describe your ideal audience for this post i.e. individuals or groups with whom you'd want to share the content. Then, describe your reasons for sharing with these individuals. For example: *"I'd share the music post with this girl in my lab because we have similar music tastes."*	text entry	
5. Think about the context in which you published the post referenced in question 3. In your own words, describe individuals or groups with whom you would avoid sharing this content. Then, describe what makes you want to avoid this communication. For example: *"I wouldn't share the post about my music with my cousin because he might judge me."*	text entry	
6. Think about the context in which you published the post. Describe in your own words, the audience you actually selected for this post (e.x. Friends, Friends of Friends, Only me, Public, Custom).	Drop-down menu with multiple choice options	• Friends • Friends of Friends • Only me • Public • Custom
7. If you're comfortable sharing, please feel free to copy and paste your post here. Please feel free to remove any overly personal/sensitive content. Again, your responses will **not** be associated with your identity.	text entry	
Block 2: 10-item IUIPC (Internet User Internet Privacy Concern Scale)		
8. Please respond to the following statements.	Seven-point scales anchored with "strongly disagree" and "strongly agree" (newly developed).	• Consumer online privacy is really a matter of consumers' right to exercise control and autonomy over decisions about how their information is collected, used, and shared. • Consumer control of personal information lies at the heart of consumer privacy. • I believe that online privacy is invaded when control is lost or unwillingly reduced as a result of a marketing transaction.
9. Please respond to the following statements.	Seven-point scales anchored with "strongly disagree" and "strongly agree" (newly developed).	• Companies seeking information online should disclose the way the data are collected, processed, and used. • A good consumer online privacy policy should have a clear and conspicuous disclosure. • It is very important to me that I am aware and knowledgeable about how my personal information will be used.
10. Please respond to the following statements.	Seven-point scales anchored with "strongly disagree" and "strongly agree" (Adapted to an internet environment (e.g., companies to online companies).).	• It usually bothers me when online companies ask me for personal information. • When online companies ask me for personal information, I sometimes think twice before providing it. • It bothers me to give personal information to so many online companies. • I'm concerned that online companies are collecting too much personal information about me.

Block 3: Covariates for IUIPC		
11. What is your gender?	multiple choice	• woman • man • non-binary • prefer not to disclose • prefer to self-describe
12. What is your age?	multiple choice	• 18-24 years old • 25-34 years old • 35-44 years old • 45-54 years old • 55-64 years old • 65 years or older
13. Education	multiple choice	• some school, no degree • high school graduate • some college, no degree • bachelor's degree • master's degree • professional degree • doctorate degree
14. Internet experience	multiple choice	• less than a year • 1–less than 2 years • 2–less than 3 years • 3–less than 4 years • 4–less than 5 years • 5–less than 6 years • 6–less than 7 years • more than 7 years
15. Some websites ask for you to register with the site by providing personal information. When asked for such information, what percent of the time do you falsify the information?	multiple choice	• I have never falsified information • under 25% of the time • 26%–50% of the time • 51%–75% of the time • over 75% of the time
16. How frequently have you personally been the victim of what you felt was an improper invasion of privacy?	slider scale	1 = very infrequently; 7 = very frequently
17. How much have you heard or read during the last year about the use and potential misuse of the information collected from the Internet?	slider scale	1 = not at all; 7 = very much

A.2 Scenario descriptions for RQ2

Descriptions of posting scenarios that were shown to participants during the user study.

Example scenario 2. Eric is a 22 year-old graduate student at [XXX]. He has been stressed out about his academic performance and the upcoming final exams. Last semester, he joined a support group on campus that aims to help students manage their anxiety, and improve mental well-being by facilitating peer support networks and study groups. Eric wants to post on Facebook about a new mindfulness app that he found very helpful in managing stress. He thinks such a tool will be helpful not only to the members of the support group but to the larger grad student population. However, he doesn't want his family or professors, who are his Facebook friends, to know about the app and his experience with anxiety.

Example scenario 3. Deepa is interested in fitness and outdoor activities. She has been training for a marathon all summer. All funds raised by the runners participating in the marathon are directed towards supporting abortion rights movements. Deepa wants to use Facebook to promote her run and raise funds for the cause. She strongly supports the right-to-chose but is worried that some of her family members and friends on Facebook might disagree with her opinions. She wants to publicize her run while trying not to offend anyone's opinions or get into arguments online.

A.3 Adjectives list from the desirability scale (Benedek and Miner (2002) [7])

[Appealing, Complex, Confusing, Customizable, Desirable, Easy to use, Efficient, Empowering, Exciting, Familiar, Fast, Flexible, Fresh, Frustrating, Fun, Gets in the way, Hard to use, Inconsistent, Intimidating, Not valuable, Overwhelming, Predictable, Reliable, Relevant, Rigid, Straight-forward, Time-consuming, Time-saving, Too technical, Trustworthy, Uncontrollable, Unpredictable, Usable, Useful, Valuable]

Received January 2021; revised April 2021; accepted May 2021