"I know even if you don't tell me": Understanding Users' Privacy Preferences Regarding AI-based Inferences of Sensitive Information for Personalization

Sumit Asthana
asumit@umich.edu
University of Michigan, Ann Arbor
Ann Arbor, Michigan, USA

Jane Im
imjane@umich.edu
University of Michigan, Ann Arbor
Ann Arbor, Michigan, USA

Zhe Chen
czhe@umich.edu
University of Michigan, Ann Arbor
Ann Arbor, Michigan, USA

Nikola Banovic
nbanovic@umich.edu
University of Michigan, Ann Arbor
Ann Arbor, Michigan, USA

ABSTRACT

Personalization improves user experience by tailoring interactions relevant to each user's background and preferences. However, personalization requires information about users that platforms often collect without their awareness or their enthusiastic consent. Here, we study how the transparency of AI inferences on users' personal data affects their privacy decisions and sentiments when sharing data for personalization. We conducted two experiments where participants (N=877) answered questions about themselves for personalized public arts recommendations. Participants indicated their consent to let the system use their inferred data and explicitly provided data after awareness of inferences. Our results show that participants chose restrictive consent decisions for sensitive and incorrect inferences about them and for their answers that led to such inferences. Our findings expand existing privacy discourse to inferences and inform future directions for shaping existing consent mechanisms in light of increasingly pervasive AI inferences.

CCS CONCEPTS

• **Security and privacy** → **Social aspects of security and privacy**; • **Human-centered computing** → *Empirical studies in HCI*.

KEYWORDS

Personalization, privacy, inference, consent.

ACM Reference Format:
Sumit Asthana, Jane Im, Zhe Chen, and Nikola Banovic. 2024. "I know even if you don't tell me": Understanding Users' Privacy Preferences Regarding AI-based Inferences of Sensitive Information for Personalization. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA*. ACM, New York, NY, USA, 21 pages. https://doi.org/10.1145/3613904.3642180

CHI '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0330-0/24/05
https://doi.org/10.1145/3613904.3642180

1 INTRODUCTION

Personalized interactions based on data that users provide with voluntary and informed consent [45, 58] can improve user experience by presenting relevant information tailored to individual interests, while giving people agency over their data [15]. This is beneficial for end users; personalization through enthusiastic consent [45] allows them to use online services without resignation about their privacy or fear of misuse of their data [10, 84, 85]. In turn, this could increase value for businesses that own the online services by broadening the user base and easing concerns for existing users [52, 60, 106].

However, current data collection practices for personalization remain at odds with obtaining voluntary and enthusiastic consent from end-users. Online services collect data required for personalization (e.g., user demographics, their interests) invisibly from their users when they engage in online activities like making purchases, surfing social media, engaging with posts, or explicitly by directly asking them questions. They then make "AI inferences"—they aggregate attributes across different users whose preferences they know and use Artificial Intelligence (AI) algorithms to infer the interests of other similar users (Figure 1). Recently, data aggregation across online services and increased accuracy of AI algorithms [14, 33, 121] has enabled broad inferences about end-users' demographics and interests, often without their knowledge [10, 37].

Thus, AI inferences make already opaque data collection and usage practices even less transparent and user consent even less informed [43]. AI inferences pose risks of negative consequences for individuals [19, 30, 48, 82] (e.g., increased health insurance premiums if insurance agencies monitor consumers' food purchasing behaviors [8]), especially when the inferences are wrong [84] (e.g., using inferences on sensitive user attributes from poorly tested AI models [29]). Yet, users have limited awareness of such inferences [22, 113, 121]. When users do become aware of such inferences (e.g., by seeing sensitive ads [83, 84]), behaviors that they adopt to safeguard their privacy are often misinformed [50] (e.g., changing browser settings, which has little effect on advertising data collection). Despite public opinions [107] that reject the use of AI inferences, users still appreciate the benefits of personalization that such inferences enable [54].

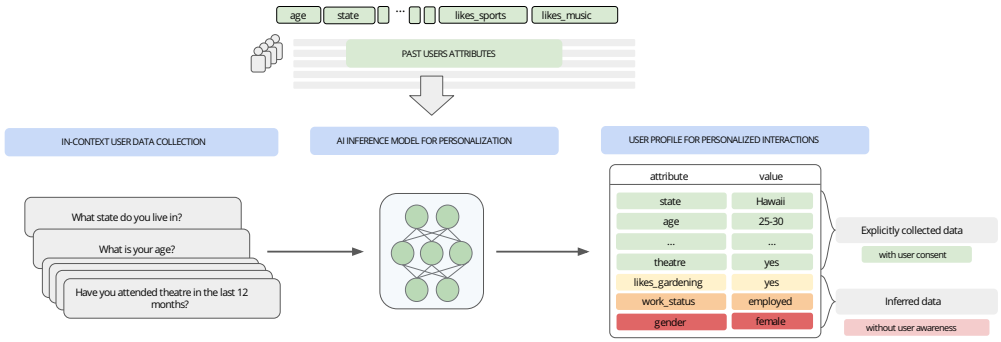


Figure 1: The figure illustrates an AI system collecting user information for personalization. The collected user information also allows inference for other user attributes that can be used for personalization or for showing ads.

Although existing research [85, 114] has called for inference transparency as a first step towards informed consent, such research primarily relied on people's retrospective opinions about their inferred data through platform-provided controls, such as privacy dashboards [115] or ad explanations [85, 114]. This precludes studying the effects of different types of inferences generated by complex "black-box" AI algorithms that the public has no insights into [46]. In particular, it precludes studying how users make *in situ* consent decisions for inferences (including data used for such inferences) made about them with varying sensitivity and accuracy [36]. Such understanding is essential to inform concrete approaches to building consent mechanisms for inferred data [10] and identify gaps to educate users for privacy-preserving online behaviors [117].

In this work, we study users' privacy behaviors and sentiments once they become aware of AI inferences generated from the data that they provided to a system for personalization. We answer the following research questions:

- RQ1: How do users *consent to the use of AI inferences* about them from the data that they explicitly provide for personalization?
- RQ2: How do users' *consent to the use of explicitly provided data* vary after becoming aware of AI inferences of varying sensitivity, and how does it compare to users' consent without the awareness of any inferences?
- RQ3: What are users' perceived informedness and sentiment towards the system's data practices after becoming aware of AI inferences of varying sensitivity, and how do they compare with perceived informedness and sentiment in the absence of inferences?
- RQ4: How do users' consent decisions for the use of the same type of personal attributes vary when they *explicitly provide* them versus when AI *infers* the attributes?

We conducted two quantitative experiments, with 877 participants across both experiments, to answer the above research questions. In both experiments, participants first answered questions about themselves to get personalized recommendations from a hypothetical personalized public arts recommendation system. The recommender system used participants' answers to questions about their public arts preferences (e.g., about dance, music, theater) to infer unknown attributes about them at three privacy sensitivity levels based on existing personal data categorizations [10, 66]: 1) attributes that described their public arts preferences, 2) attributes relevant for online behavioral advertising (e.g., household income), and 3) protected attributes (e.g., race) [4, 10]. In the first experiment, we asked participants (N=333) to consent to the use of the attributes that the recommendation system inferred about them from their answers. In the second experiment, we asked participants (N=544) to consent to use their answer after the recommendation system showed them an inference of an attribute about them from one of the three categories or consent to use their answer without seeing any inferences. In both experiments, we asked participants to indicate consent on a 4-point scale: 1) *no consent*, 2) *consent to use for personalized public arts recommendations*, 3) *consent to use for ads within the platform*, and 4) *consent to use for ads outside the platform*. Each successive level included consent for previous levels and indicated broader data use (similar to website cookie consent interfaces [53]).

Our results highlight that participants took a nuanced approach toward consent decisions in the presence of inferences. Participants were less likely to give their consent for usage of the most sensitive inferences (e.g., citizenship status) beyond personalization. They were also likely to give lower consent decisions for the use of their answers after they became aware of AI inferences generated from their answers. Moreover, participants who saw incorrect inferences about them from the recommendation system chose lower consent decisions, both for the use of such inferences and their answers compared to other participants. Participants also gave lower consent decisions for the same type of attribute when the system inferred it versus when participants explicitly provided it as an answer. We also found preliminary evidence that transparency of most sensitive inferences increases the system's appeal about its data practices more than other inferences. Participants' differential decisions and attitudes towards inferences and their answers after awareness of inferences highlight the importance of rethinking companies' data practices by centering users' privacy awareness and needs around AI interfaces. Our work informs the design of future interfaces that ask users for their informed, enthusiastic consent about inferences that will enable users to benefit from AI personalization without concern about their privacy [36].

2 RELATED WORK

Here, we first provide a broad overview of the model of notice and choice—which is the dominant way for organizations in many countries to obtain users' consent [79], as well as existing research on its shortcomings. Then, we review studies on people's privacy perceptions and behaviors regarding inferences of personal information. Lastly, we discuss work on protected attributes, which are types of sensitive demographic information that are closely related to privacy concerns regarding the rise of AI-driven applications.

2.1 Notice and Choice Model and Users' Privacy Concerns with Online Services

The way companies and other organizations obtain people's consent regarding data privacy is centered around a paradigm called "notice and choice" [79]. This paradigm emphasizes giving individuals information (notice) and control mechanisms for deciding how one's information is collected and used (choice) [79]. While the initial aspiration of the model was to enhance privacy protection for users, scholars have criticized that it has failed in practice [13, 62, 76, 99]. For example, users cannot keep up with an overwhelming amount of complex information in order to make an informed privacy decision [88, 93]. Also, studies have shown that people do not read privacy policies and terms of service [26, 47, 79]. Privacy choices provided by companies are often meaningless and have low usability [13, 25]. Many users digitally resign to share their data in order to use a service and cannot indicate granular levels of consent for different kinds of data or interactions [92, 93].

Users are increasingly feeling concerned about online services' data privacy practices [54, 56, 70]. Such sentiment is related to a lack of transparency about the collection and usage of data [24, 103], as well as a lack of control regarding data deletion [39]. Although internet users may be largely aware of companies collecting personal data for their services, such as for online advertisement and personalization [23], they remain largely unaware of the extent of companies' tracking practices across platforms [109, 114, 121]. However, such increasing awareness and concerns do not mean users understand how AI inferences work.

Although users may express concerns for privacy in self-reported evaluations [63, 86], their online behaviors often contradict such concerns, leading to a privacy-paradox [2, 77, 77, 101]. For example, social media users may express desire for privacy, but may engage in sharing behavior going against their privacy preferences [44, 64]. However, recent critiques [1, 32] argued that the "paradox" is an oversimplification of user behaviors, which could arise from a lack of adequately designed privacy consent mechanisms to support such nuanced users' online behaviors [49]. Companies typically do not provide consent mechanisms that ask for users' preferences on both the data used for inferences and inferred information. However, the context and design of consent are as important as asking for consent itself for data use [58, 75]. Im et al. [45] provide a framework to think about problems with existing consent mechanisms as violating one or more of the following dimensions of consent: *voluntary, informed, reversible, specific*, and, *unburdensome*. For example, large-scale passive data collection through privacy notices makes consent burdensome and inadequate to keep users informed.

2.2 Users' Privacy Perceptions Regarding Inference of Personal Information

Users can develop an uninformed understanding of AI inferences about them due to limited knowledge of the service's data practices and how inferences are generated [113, 121]. This limited understanding of AI inferences and awareness can lead to misinformed online behaviors that may not preserve privacy and also hurt personalization experience [54]. For example, in self-reported measures, users have indicated holding back information (e.g., not posting on social media), strategically sharing it [73, 80] or even avoiding the use of the service (e.g., not purchasing a product due to privacy concerns [9]). Users may feel resignation towards their privacy, hoping that online services will act in their best interest [56].

Studies that probed users about their perceptions of inferences through retrospective surveys and dashboards [117] suggest a rejection of sensitive inferences like location or demographic inferences for third-party use. When users interact with their inferred information to make sense of it (e.g., in the form of realistic ad profiles) [9, 116], it can lead to more informed awareness about the privacy implications of their behaviors and increased interest in privacy-protective actions [116]. However, users' perceptions regarding online inferences elicited through interviews, diary studies, and surveys in contexts such as online behavior advertising [22, 85, 121], voice assistants [56, 65, 111], and chatbots [28, 61] indicate how users make sense of inferences but do not tell us about their privacy decisions for their data with AI inferences.

2.3 Collecting and Inferring Users' Protected Attributes for AI-Driven Personalization

With the rise in AI-driven applications, there has been an increasing discourse around the use of "protected attributes"—demographic information (e.g., race, religion, gender) that are sensitive in nature and can be abused for discrimination [72, 89]. Major tech companies are facing increasing public scrutiny as audits have uncovered algorithmic discrimination [3, 18, 102, 110]. For example, researchers found that Facebook's AI algorithms delivered ads for employment and housing that were skewed based on gender and race—even with neutral ad targeting parameters [3]. Despite the sensitive nature of protected attributes, fully restricting their collection across all kinds of services may not be the answer [10, 89, 120]. Instead, access to such data is needed in the first place in order to mitigate biases [120]. Furthermore, excluding protected attributes does not guarantee the elimination of bias because other attributes could be correlated with them [40, 105].

Yet, in the U.S., the anti-discrimination doctrine has a preference for restricting the existence of protected attributes in data and algorithms [120]. However, a regulation that guides data collection and usage practices of protected attributes is ambiguous and not straightforward for industry practitioners [5]. This has led to inconsistent practices in how businesses decide what to infer and how to use protected attributes in AI algorithms, with U.S. civil rights laws affording major tech companies special legal protections from liability arising from collection and use of protected attributes [10]. However, designing consent mechanisms for allowing companies to infer sensitive information is fundamentally a much more complex issue than collecting it directly.

3 METHOD FOR STUDYING AND UNDERSTANDING PRIVACY PREFERENCES

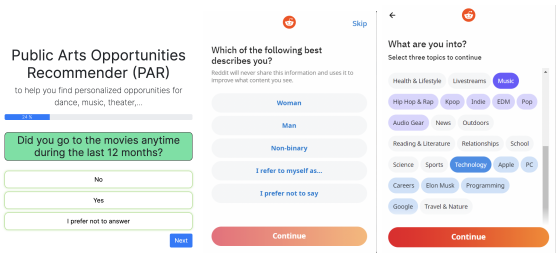
This work aims to understand *how* users make in-situ privacy decisions for the use of different kinds of inferences about them (including those about protected attributes) and for the use of their explicitly provided data after becoming aware of different inferences. Such understanding can help us carve out nuanced approaches to seeking users' consent to use explicitly provided data and their inferences.

To answer our research questions, we conducted two lab experiments in which participants interacted with a hypothetical Public Arts opportunities Recommender (PAR) system, which we developed for our method. PAR simulates a "cold-start recommendation scenario" where a recommendation system asks questions to know their users when they first sign up on the platform (e.g., streaming service, insurance quote personalization, shopping recommendations). Asking questions to users upfront provides recommender systems with explicit information about them and is widely used for personalization by existing online services [34]. Figure 2 shows an example from reddit. PAR focuses on recommendations for public arts opportunities (e.g., dance opportunities, music events, theater).

We used this experimental setup because the data collected by online services is not accessible to the public. The cold-start scenario provides a good first step for studying privacy decisions and sentiment regarding inferred data because users explicitly provide information to the system and can easily relate to this information when providing consent decisions about the use of their data in the presence of AI inferences.

3.1 Hypothetical Recommender System

Here, we describe our hypothetical public arts opportunities recommender (PAR) system. For the purposes of our study, we only needed a subset of functionality of a real recommender system: 1) a user model that can record information about our participants and that we can use to make inferences about them, 2) a set of questions that map onto the variables in the user model, and 3) a mechanism that selects the most informative set of questions that the system can ask the participants to enable cold-start personalization and enable inference of unknown attributes about the user [7].



(a) (b) (c)

Figure 2: Screenshots of questions that systems ask their users: a) our hypothetical recommendation system asking about user preferences, b) a real-world example from Reddit asking about user demographics, and c) their preferences.

3.1.1 Modeling Users and their Preferences. Like real-world recommender systems, our hypothetical recommender system needs to keep track of user information. We represent information about the user as a Bayesian Network $\mathcal{G}$, where nodes represent a set of variables $\mathcal{U}$ that describe the user (e.g., their demographics, public arts preferences), and the edges ($\mathcal{E}$) represent conditional dependencies between those variables. Mathematically, we represent $\mathcal{U}$ as a set of random variables:

$$\mathcal{U} = \{X_1, X_2, \ldots, X_i\}$$
$$i \in \{age, gender, ..., likes\_sports, likes\_music\} \quad (1)$$

Each discrete random variable $X_i \in \mathcal{U}$ corresponds to one *attribute* about the user (e.g., age, gender, likes going to sports events). The system maintains knowledge about the user by assigning a probability distribution to each random variable. The system either infers those probability distributions or assigns an explicit value to a variable when the user answers a question about that variable.

Initially, when the system knows nothing about the user, all variables have an initial prior probability distribution. Once the user answers a question, for example, "What is your age?" with "18-25", the system sets an explicit value for that attribute to the user-provided answer. Knowing values for some variables enables the system to infer the most likely values of other variables (i.e., estimate their conditional probabilities given the user's answers) in the graph $\mathcal{G}$. For example, if the system knows that the user's age is "18-25", it can infer that the user's most likely profession (highest probability) is "student".

3.1.2 Learning user model parameters. To build a user model for public arts recommendation, we used a publicly available dataset from the Supplement of Public Arts Participation (SPPA) from the US Census Bureau's Current Population Survey (CPS) [1] [27]. The CPS dataset contains records of approximately 125,000 people from 60,000 households, which detail their demographics (e.g., age, gender), family status (e.g., marital status, children), and their interest in public arts activities (e.g., participation in performing and visual arts, literature, museums and galleries, art classes and lessons). We excluded records from people under 18 years old or where a full interview was not conducted (e.g., respondents refused to answer or answered "Don't know"). In addition to loading the data, we upsampled it based on the Census Bureau computed person weight statistical variable, which indicates how representative each sample of people in the United States is. By doing so, our final pre-processed dataset can approximate the data for all adults in the US. This resulted in 8.2M final records.

Each question in the dataset maps to one *attribute* about the user (e.g., age, employment status). Thus, each question in the CPS survey corresponds to one variable in our user model $\mathcal{U}$. We eliminated variables from CPS that were related to survey format (e.g., response mode, person number), variable flags and statistical weights, variables that are only relevant to minors (e.g., nativity of parents, employment of parents), and duplicates of the same information (e.g., same occupation based on multiple internal codes).

We then transformed each variable's possible values from the dataset to discretize them in a way that allowed us to build the

[1] https://www.census.gov/programs-surveys/cps.html

user model efficiently, such as binning for continuous variables and grouping for variables with a large number of possible answers (e.g., countries). We then used our dataset to learn the structure for $\mathcal{G}$ using the ARACNE algorithm [67] in the bnlearn package [95]. To learn the conditional probability distributions in the Bayesian network, we use the parameter estimation procedure using pgmpy [6]. Table 1 illustrates the initial probability distributions in the learned user model about a user's age and marital status.

Table 1: Normalized initial probability distributions for age and marital status after learning the user model parameters. For initial probability distributions of all questions, refer to the supplementary materials.

Question	Answer	Probability
Age	18 - 20	0.03
Age	21-40	0.33
Age	41-60	0.29
Age	60+	0.33
Marital status	Married, spouse present	0.45
Marital status	Married, spouse absent	0.01
Marital status	Separated	0.02
Marital status	Divorced	0.15
Marital status	Widowed	0.09
Marital status	Never married	0.26

3.1.3 Question wording. The original survey questions that correspond to the CPS variables were not worded in a way that a recommender system would ask them (e.g., they are worded in the third person, such as in "Person 1's age"). We modified the questions so that the system could ask them in first person (e.g., changed "Person1's age" to "What is your age?"), and made numeric answers more verbose (e.g., changing 18-25 to "18 or more but less than 25").

3.1.4 Categorizing sensitivity of questions. Not all questions in CPS are relevant or appropriate for a public arts recommendation system to ask. As described in Section 2.3, U.S. anti-discrimination doctrine has a strong preference towards restricting protected attributes (e.g., citizenship, race) in data and algorithms to prevent discrimination [120]. Further, every platform can only ask questions that are plausible to the context of the service even if other questions may be more informative (e.g., a digital music online service can ask its users about their music preferences or questions that are plausible in the context of music, such as their age).

Thus, we conducted multiple rounds of annotation on the entire set of questions to categorize them based on their appropriateness of being asked by a realistic public arts recommender system. The first and second authors read a set of randomly selected 26 questions (25% of the original 104 questions) and collaboratively drafted a codebook, where the criteria were: 1) whether the question is related to public arts, 2) whether it is socially acceptable for a system that recommends content related to public arts, and 3) whether the answer to the question can be used to recommend ads, 4) Whether the attribute belongs to a protected class [4].

We defined a "socially acceptable" question as one that: 1) is normatively asked by services even if it is not directly related to the application (e.g., asking about gender is widespread across online services, see Figure 2-b), and 2) does *not* mention any socioeconomic status (e.g., type of work, wage, education). We erred towards not including a question if there is any slight possibility it could offend a user. An example of offending the user is asking about someone's disability or their wage. We used existing research [14, 70, 117] to determine an attribute's suitability for ads and found that most attributes fell within this bucket. E.g., asking someone's wage is not socially acceptable, but this knowledge helps companies recommend more economically relevant products. Therefore, we first identify whether an attribute is related to public arts or protected.

Next, each author independently coded another set of 26 randomly selected questions. Then, the two authors met to discuss and resolve disagreements. The inter-rater reliability (IRR, Cohen's κ) for the three categories was 0.83, 0.65, and 1. The disagreements were mostly due to the lack of preciseness on what "socially acceptable" questions meant. Based on the new codebook, the first and second authors independently annotated another set of randomly selected 33 questions. This time, each had high IRR: 1) whether the question is related to public arts (Cohen's κ=0.94), 2) whether the question is socially acceptable (Cohen's κ=1), and 3) whether the answer to the question could be used to recommend any ads (Cohen's κ=1). As the IRR was high for all three, the first and second authors independently re-annotated the remaining 19 questions based on the final codebook. Then, the authors met to resolve the remaining minor disagreements.

Based on the final annotation, we assigned each question to one of the four categories: 1) *Protected* (sensitive attributions, such as race, whose usage in algorithms is restricted by the U.S. law to minimize potential discrimination) [4, 105], 2) *Public Arts* (questions that explicitly ask about public arts participation), 3) *Ads* (questions that do not explicitly ask about public arts participation, but can be used to target ads, such as age, employment status [3, 100]), and 4) *Implicit* (questions that neither explicitly ask about public arts nor are useful to target ads). We also excluded any questions related to disability and survey methods. The final set included 88 questions, with each question assigned to one of the four categories of user attributes. We provide the full list of included and excluded questions along with the annotation details in the supplementary materials. Table 2 provides example questions and their possible answers from the final list that we used in our experimental setup.

3.1.5 Selecting which questions to ask to learn about the user. A recommender system can only ask a limited set of questions to the users before overwhelming them and hurting the user experience. Thus, the system's goal is to select a subset of variables that maximizes the knowledge about all user variables relevant to the application and business [34].

To select questions that maximize knowledge about the user for public arts and ads, we adapted an existing question selection method [7] that allows question selection to maximize knowledge about specified user attributes. We use the method to maximize knowledge about "public arts" and "ads" attributes about the user. However, we restrict the system to ask "public arts" and "implicit" questions only as realistic recommender systems ask about user attributes relevant to the application but infer attributes relevant

Table 2: Examples of selected questions with corresponding possible answers separated by the "|" symbol.

Question	Answer Options
Have the children in your household experienced cultural events or art venues?	Yes|No
What is your citizenship status?	US citizen|Naturalized citizen|Not a citizen
What is the age of your eldest child?	0-5|6-10|...|90+
How many members are present in your family?	0|1|2|3|4-7|7+

to the business. Table 3 summarizes this distinction. The output of this step is a sequence of questions to ask users that maximizes the information about all possible user attributes (except "protected") and enables the system to infer the remaining attributes that it did not ask the user about.

Table 3: Categories of user attributes that the system can ask, learn about, and infer.

Category	Can ask	Can learn about	Can infer
Arts attributes	yes	yes	yes
Implicit attributes	yes	yes	yes
Ads attributes	no	yes	yes
Protected attributes	no	no	yes

3.1.6 Inferring unknown user attributes using collected data. After the system is done asking a small fixed number of questions to know as much about the user as possible, the system has a partially observed user model.

$$\mathcal{U} = \mathcal{A}_s \cup \mathcal{U} \setminus \mathcal{A}_s \tag{2}$$

$$\mathcal{A}_s = \{X_1^a = x_1, X_2^a = x_2 \ldots X_i^a = x_i\} \tag{3}$$

$$\mathcal{U} \setminus \mathcal{A}_s = \{X_1^r, X_2^r \ldots X_j^r\} \tag{4}$$

User $\mathcal{U}$ is now described by a union of the variables that the system asked the user X^a and has answers to (asked and answered) and the variables that the system does not have an answer to X^r (remaining). The lowercase x_i's represent assigned values for the variable X_i obtained from the user's answer. Values for X_j^r's are still unknown.

Based on the answers provided by the user, the most likely values of the remaining attributes can be *inferred* by taking the argmax of the conditional probability distribution for the remaining variables X_j^r conditioned on the answers obtained from the user $\mathcal{A}_s$.

$$Answer(X_j^r) = argmax_{X_j^r \in \{x_1, x_2 \ldots x_k\}} P(X_j^r \mid \mathcal{A}_s) \tag{5}$$

$x_1, x_2 \ldots x_k$ are the k possible values that the variable X_i^r can take.

3.1.7 Identifying the most informative user answers contributing to each inference. Although all answers provided by the user contribute to inferences of unknown attributes, they are not all equally informative for a specific inference. We identify the *most informative answer for the inference* (X_i^a) as one that causes the maximum difference in the conditional entropy of the inferred user attribute:

$$X^a = argmax_{X_i^a \in \mathcal{A}_s} P(X_j^r \mid \mathcal{A}_s) - P(X_j^r) \tag{6}$$

Equation 6 selects the answer that provides the most information about an inference to study how users consent to use that answer when they learn that it was used to generate an inference.

3.1.8 Cold-start Personalization User Interface. To administer the study using our system, we developed a user-facing web interface called "Public Arts opportunities Recommender (PAR)" to recommend "personalized" opportunities for public arts (dance, music, theater) to its users. The system asks ten questions to maximize knowledge about the user. After the user answers the questions, the system uses these answers $\mathcal{A}_s$ to generate inferences about them using the user model described in Section 3.1.1.

3.2 Consent Operationalization

Personalized systems can ask users if they consent to 1) systems generating and using inferences about them (i.e., asking before generating the inferences) or 2) using inferences after the systems have already generated them. Our work explores the latter. In addition to studying consent to generate inferences, this also allowed us to study how users make privacy decisions after being made aware of correct and incorrect inferences of varying sensitivities, which the online services are already using opaquely [85]. Our approach did not preclude us from exploring users' refusal of inferences (e.g., indicating they do not want the inferences to be used for anything).

To measure consent decisions, we take inspiration from the idea of consent levels operationalized by cookie consent interfaces that are mandated by the EU's ePrivacy Directive (EPD), General Data Protection Regulation (GDPR), and the California Consumer Privacy Act (CCPA) [31, 53, 78, 108]. Specifically, the interfaces meet the requirements of the Transparency and Consent Framework (TCF) developed by IAB Europe to comply with GDPR requirements. At each consent level in the cookie consent interface, systems ask users' preferences for more data collection for broader uses. However, the regulation leaves room for the design choices for seeking consent, and thus, several design variations have come up in the notice and choice realm [12].

We modeled consent as successive levels of broader data use on a 4-point scale indicating what the information will be used for: 1) *no consent*, 2) *personalized public arts recommendations*, 3) *ads within the platform*, and 4) *ads outside the platform*. Each consent level includes consent for that level and all previous levels. For example, the highest level of consent (*ads outside the platform*) implies that users consent to the use of their data for all of the previous levels. The lowest level of consent (*no consent*) explicitly states that users provide no consent for the use of their data. This formulation is in line with existing consent practices [11, 36, 38] while allowing our study participants to express consent for the use of inferences.

3.3 Overview of User Study Experiments

Using our method and the study software, we explore our four research questions. We split our study into two experiments: 1) an experiment to study the *consent decisions for the use of users' inferred attributes* that PAR generates from their answers ($\mathcal{A}_s$) (RQ1) and 2) an experiment to study the *consent decisions for the use of users' answers and their sentiment of PAR's data practices* after PAR makes them aware of inferences generated from their answers (RQ2 & RQ3). To compare how users' consent decisions vary for the same type of attribute when PAR asks it versus when PAR infers it (RQ4), we analyze consent for using arts inferences from Experiment 1 and consent for using answers to arts attributes from Experiment 2. We conducted both experiments on the Prolific[2] platform.

3.4 Pilot User Studies

To assess the clarity and interpretation of the consent interface, we conducted pilot studies with four participants (three graduate students and one professional; none were closely associated with the privacy research area). No participant had prior knowledge of the study. In addition to our successive 4-point scale design, we showed participants an alternative design with three consent levels corresponding to using their data for 1) *personalized public arts recommendations*, 2) *ads within the platform*, and 3) *ads outside the platform*, with an option to select zero or more of them at the same time. We asked for participants' understanding of the consent through think-aloud verbalizations [21].

All participants correctly understood how the two designs asked them to indicate their consent. However, three out of four participants indicated that it was less cumbersome to choose consent using the successive-level scale because they understood that broader use cases (e.g., ads) would encompass narrower use cases (e.g., personalization). Thus, we decided to keep the 4-point successive-level scale interface for seeking consent. We also asked participants to provide their feedback on the clarity of questions that PAR asked for personalization, as well as the clarity of the inferences and design of the consent page. We received comments on using colors to explicitly indicate whether PAR is asking consent for the use of users' inferred data or for the use of their answers.

We also piloted the interface for both experiments with an additional 50 participants on Prolific to evaluate if PAR generated and displayed the inferences correctly and that all parts of the system, such as question-asking, consent for data, and post-experiment survey, worked correctly. We asked for pilot Prolific participants' feedback via an open-ended question for feedback at the end of the experiment but did not receive any comments. We compensated all pilot participants $15 for piloting the study. We provide more details on the pilot studies in the supplementary materials.

3.5 Experiment 1: Privacy Preferences for Inferred Personal Attributes

In this experiment, we answer our first research question (RQ1): How do users *consent to the use of AI inferences* about them, generated from the data that they explicitly provide for personalization? To answer RQ1, we study the effect of category and correctness of the AI-generated inference on users' consent level for the usage of the inference by the system. Thus, we have two hypotheses:

- H1.1: Users will choose *lower consent levels* for the usage of protected inferences, compared to inferences generated in the arts and ads categories.
- H1.2: Users will choose *lower consent levels* for the usage of incorrect inferences about them compared to correct inferences across all categories.

The reasoning behind H1.1 is that prior studies have demonstrated that users tend to reject sensitive inferences when probed about them in retrospective evaluations [54]. By understanding how users consent to the use of inferences, we aim to move towards personalization using inferred data with the user's consent. In H1.2, we reasoned participants would dislike a system having incorrect information about them as it could lead to irrelevant content recommendations, system actions, or unexpected consequences in the case of a data breach [85].

3.5.1 Study Design. In our study design, we varied the category of the inferred attribute ($Category$), which could take on three values: 1) $arts$, 2) ads, and 3) $protected$. We asked each participant to consent to the use of an inference from each category, making this a within-subjects factor. Because inferences can be either correct or incorrect, we also consider the correctness of the inference ($Correctness$), which we assessed using a post-experiment survey. Thus, our design is ($Category \times Correctness$), with $Correctness$ being a nested and unbalanced factor [55].

We measured participants' consent level $Consent_I$ for the use of each inference that we made using our operationalization of consent as a 4-point successive level scale from Section 3.2. We randomized the order of inference categories using Latin square sequencing to avoid bias due to the order in which the system asks for consent to use the inferences [35]. Similar to existing research, we also measured participants' overall privacy concerns on a 5-point Likert scale by asking for their ratings along three dimensions: 1) $Invasive$, their overall concern about the system invading their privacy [51], 2) $Misuse$, their concern about their private information that the system collected being misused [51, 104], 3) $Comfort$, how comfortable they feel about the system learning information about them to recommend content [12].

3.5.2 Tasks and Procedures. We conducted this experiment on the Prolific platform. The participants signed up for our study by selecting it from a list of available Prolific studies. After signing up for our study, they were redirected to our study web page. The landing page showed the participants our study consent form, which included instructions explaining the details of the experiment and that it did not pose any risk or discomfort other than evaluating questions about their personal data. Only those who consented were allowed to participate and proceed.

The next web page then instructed the participants to imagine that they have "just signed up to a new system that recommends *personalized* public arts opportunities (e.g., theatre, dance, music) and needs to learn about them to provide personalized recommendations". To further ground the participants, we mentioned existing online services (e.g., Spotify) that ask users questions to learn their

[2]https://prolific.co

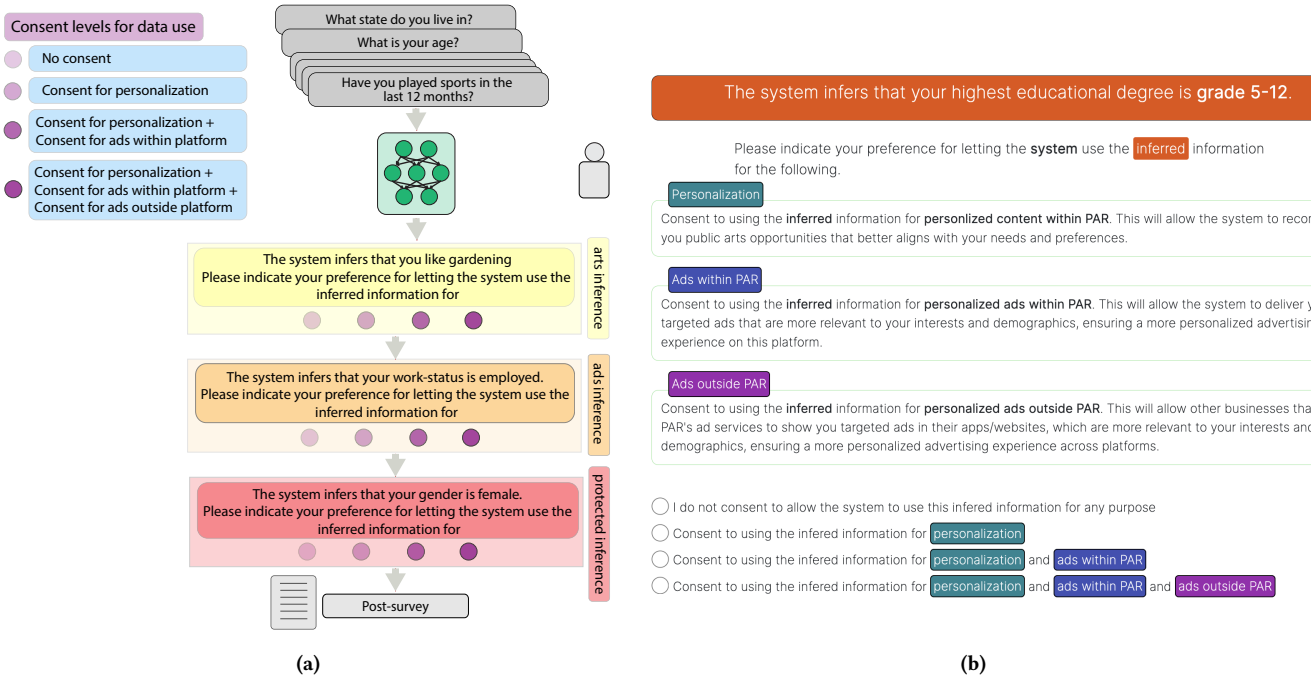


Figure 3: Experiment 1: a) experimental setup, and b) a screenshot of the inference consent page.

preferences for personalized content recommendations. Next, the study web page presented the PAR interface to the participants.

PAR asked participants ten questions from the *arts* and *implicit* question categories based on our question selection algorithm (Section 3.1.1). It displayed one question (along with its possible answers) per page (Figure 2a). In addition to the possible answers, for each question, participants could select "I prefer not to answer" if they decided not to answer the question for any reason. If the participant selected this option, the system would not receive any information about the participant.

After the participants answered ten questions about them, PAR made inferences about all of the remaining attributes that it did not ask or did not get an answer about. PAR then selected one attribute per inference *Category* and asked participants to choose their *Consent* level for the use of each of them using our study setup. PAR selected the inference with the least entropy, which implies the algorithm had the highest confidence in that inference for that category.

After participants indicated their *Consent* levels for all categories, the system displayed a "debrief" page. The page indicated to the participants that PAR is done asking questions and that subsequent questions are from the research team. The debrief also reminded them that their correct responses to the subsequent questions were important for the research study. After the debrief, we administered the post-experiment survey asking about the correctness of the inferences and questions about their privacy concerns described in the previous section, followed by demographic questions about their age, gender, highest education, occupation, and industry of work. We compensated the participants at the rate of $15/hr for their time.

3.5.3 Analysis. We analyzed *Consent* using two-way mixed ANOVA (*Category* × *Correctness*). Note that *Correctness* was an unbalanced factor as we cannot fully control for it in our experimental design. To account for its moderate imbalance, we use it in the random slope and use Type III ANOVA that mitigates the effect of unbalanced factors [42, 97]. Because the objective *Consent* was not normally distributed, we performed Align Rank Transform (ART) [118] before running ANOVA tests and performed post-hoc pairwise analyses using ART-c [20] with Holm-Bonferroni corrections. Our *a priori* power analysis ($\alpha = 0.05, 1 - \beta = 0.95$) estimated that the experiment required 333 participants.

3.5.4 Participants. We recruited 333 participants from the Prolific platform. We balanced our recruitment pool on the available gender representation in Prolific to ensure fair representation of preferences. In this experiment, about 48% participants self-identified as male, 49% self-identified as female, and 1.5% self-identified as non-binary. We had a fair representation for most age groups except 65+, with the highest representation from 25-34 (38%) and the second highest from 35-44 (23%). Many of our participants had "4-year college degrees" (40%) followed by "Some college" (21%). We provide the full demographic breakdown in supplementary materials.

3.6 Experiment 2: Privacy Preferences for Explicitly Provided Personal Attributes Used in Inferences

In this experiment, we answer RQ2, RQ3, and RQ4. To answer RQ2 (How do users' consent to the use of explicitly provided data vary after becoming aware of AI inferences of varying sensitivity,

and how does it compare to users' consent decisions without the awareness of any inference?), we hypothesized that:

H2.1: Users who become aware of inferences will choose lower consent levels for the use of their answers compared to users who were not made aware of any inferences.

H2.2: Users who become aware of protected and incorrect inferences will choose lower consent levels for the use of their answers compared to users who choose consent levels for their answers with awareness of other inferences.

To answer RQ3 (What are users' perceived informedness and their sentiment towards the system's data practices after becoming aware of AI inferences of varying sensitivity and without the awareness of any inference?), we build upon prior research [54], which indicated that inferences play a role in shaping user perception of a system's data practices. Thus, we hypothesize:

H3: Users who provide consent for the use of their answers after being made aware of inferences will have more positive sentiments about the system's data practices and feelings of informedness than users without the awareness of any inference.

To answer RQ4 (How do users' consent decisions vary for the use of the same type of personal attributes when they explicitly provide the data versus when AI infers the attributes?), we hypothesize:

H4: Users will choose higher consent levels for the use of their attributes if they explicitly answer them than when those attributes are inferred by the system.

No prior work has considered how consent for the use of the same attribute will be different when it is inferred versus when users explicitly provide it as an answer.

3.6.1 Study design. Here, we varied the category of the inferred attribute (*Category*) that PAR made participants aware of before asking for consent to *use* one of their explicitly provided answers. *Category* could be: 1) *none*, 2) *arts*, 3) *ads*, and 4) *protected*. We randomly assigned each participant to one *Category*, which determined the category of inference that PAR made the participant aware of. PAR did not make any of the participants assigned to the *none Category* aware of any inferences.

After each participant had answered the questions, PAR identified the inference with the least entropy from the category they were assigned to. For participants assigned to the *none* category, PAR identified the inference with the least entropy from any of the other three categories. For all participants, PAR then identified the *most informative answer for the inference* (using Equation 6) and showed it to the participant (Figure 4b). This is one of the answers that the participants provided to the system in the personalization phase. If PAR assigned the participant to one of the three inference categories (*arts*, *ads*, or *protected*) it also showed the least entropy inference to the participant.

Finally, PAR asked participants to provide their consent to use the *most informative answer for the inference* either after seeing the inference or without inference awareness, depending on their assigned *Category*. Since each participant is only shown answers that contributed to an inference from one of the categories, *Category* is a between-subjects factor. Because inferences can be either correct or incorrect, we consider the correctness of the inference (*Correctness*), which we assessed using a post-experiment survey. Thus, we used a nested design [55] ($Category \times Correctness$), with *Category* a between-subjects factor and *Correctness* a nested factor (which was not available for $Category = none$).

We measured participant's consent level ($Consent_A$) for using their *most informative answer for the inference* (H2.1 and H2.2). We measured $Consent_A$ using our operationalization of consent as a 4-point successive level scale from Section 3.2.

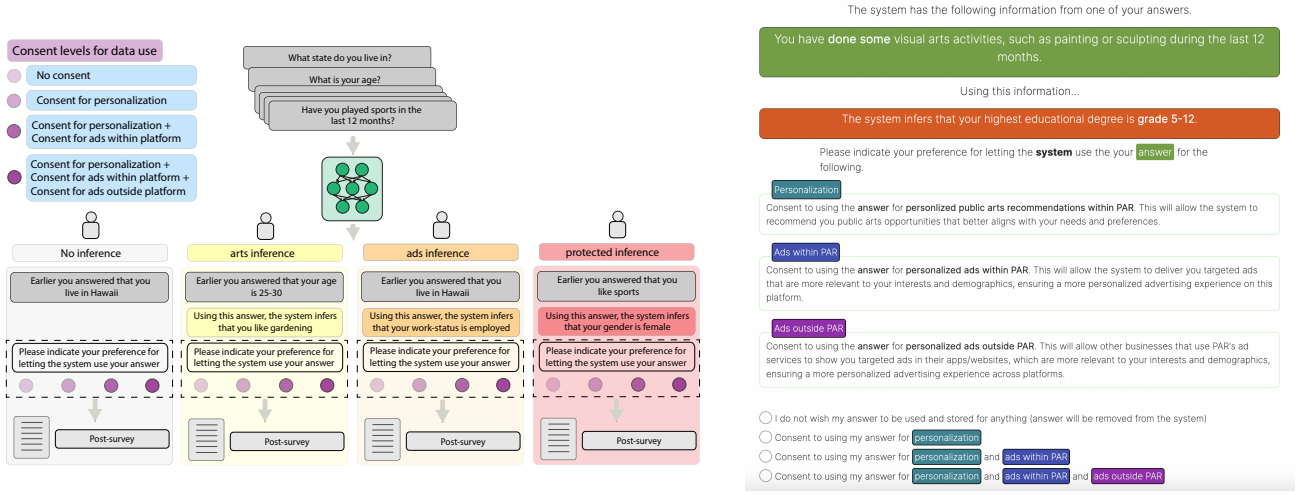
We tested H3 in a post-experiment survey. On a 5-point Likert scale, we measured participants': 1) *Sentiment*—their overall feeling when the system asked their consent to use their answer, and 2) *Informedness*—the extent to which they felt informed when the system asked for their consent to use their answer [12]. Similar to Experiment 1, we also measured participants' overall privacy concerns on a 5-point Likert scale by asking for their ratings along three dimensions: 1) *Invasive*, their overall concern about the system invading their privacy [51], 2) *Misuse*, their concern about their private information that we collected being misused [51, 104], 3) *Comfort*, how comfortable they feel about the system learning information about them to recommend content [12].

To test H4, we took $Consent_I$ for the use of inferred *arts* attributes for participants from Experiment 1 and called it *Inferred*. We took $Consent_A$ for participants who are part of category *none* and who provided consent for the use of an *arts* answer in Experiment 2 and called it *Explicit*. From Experiment 1, we only used *Inferred* from participants who saw the *arts* inference first so that their consent is not biased by other inferences. Thus, we can compare consent levels to use the same type of attribute (*arts*) when it is inferred by PAR (Experiment 1) and when it is explicitly provided by participants to PAR as an answer (Experiment 2).

3.6.2 Tasks and Procedures. We conducted this experiment on the Prolific platform, too. This experiment followed the same signup and consent procedure from Experiment 1 (Section 3.5.2). Once participants read the instructions and consented to our study, PAR assigned them to one of the four inference categories (*Category*). PAR instructed the participants to imagine the same hypothetical context. PAR then asked participants ten questions from the *arts* and *implicit* question categories in the same way as in Experiment 1, and inferred their remaining personal attributes.

PAR then selected the inference with the least entropy based on the participant's assigned *Category*. For participants in *Category* other than *none*, it showed the inference; it showed no inference when *Category* was *none*. PAR then showed the participants the *most informative answer for the inference* (that it identified using Eq 6 for the inference) and asked them to provide consent for the use of their answer ($Consent_A$) using our study setup (Figure 4b).

After participants indicated their consent level for their answer, the system displayed a "debrief" page indicating the end of interaction with PAR. The debrief also indicated that subsequent questions were from the research team and reminded them that correct responses to the post-experiment questions are important for the research study. After the debrief, we administered the post-experiment survey asking about inference correctness and participants' privacy concerns described in the previous section, followed by demographic questions about their age, gender, highest education, occupation, and industry of work.



(a) Experiment 2 flow diagram **(b) Experiment 2 consent page**

Figure 4: Experiment 2: a) experimental setup, and b) the consent page asking users for consent to use their answers.

3.6.3 Analysis. To test H2.1, we analyzed $Consent_A$ using a one-way ANOVA, with *Category* as the independent variable. Note that we do not consider correctness when *Category* is *none* due to the absence of an inference. To test H2.2, we analyzed whether there is any difference in the consent decisions of participants ($Consent_A$) for the three inference categories (*arts*, *ads*, and *protected*) and correctness (*Correctness*) using a two-way mixed ANOVA (*Category*× *Correctness*). We experimentally balance the inference *Category*, but cannot control for *Correctness*. To account for the moderate imbalance in correctness, we used Type III ANOVA that mitigates the effect of unbalanced factors [42, 97]. To test H3, we analyzed *Sentiment* and *Informedness* using one-way ANOVA (*Category*).

To test H4, we compare consent to use *Inferred* arts attributes (Experiment 1) with consent to use explicitly answered arts attributes (*Explicit*, Experiment 2). Because this analysis has only one independent variable with two possible values, we ran Welch's two-sample t-test on the consent decisions *Inferred* and *Explicit*.

None of our dependent variables in any of our ANOVA tests was normally distributed; therefore, we performed Align Rank Transform (ART) [118] before running the ANOVA tests and performed post-hoc pairwise analyses using ART-c [20] with Holm-Bonferroni corrections. Our *a priori* power analysis ($\alpha = 0.05$, $1 - \beta = 0.95$) estimated that the experiment required 540 participants.

3.6.4 Participants. We recruited 544 participants from Prolific and compensated them at $15/hr for their time. We balanced our recruitment pool on the available gender representation provided by prolific to ensure fair representation of preferences. In this study, about 48% participants self-identified as male, 47% self-identified as female, and 3% self-identified as non-binary. We had a fair representation for most age groups except 65+, with the highest representation from 25-34 (42%) and the second highest from 35-44 (21%). Many of our participants had "4-year college degrees" (38%), followed by "Some college" (23%). We provide the full demographic breakdown in the supplementary materials.

3.7 Summary of the Experimental Design

Experiment 1 is within-subjects with inference *Category* as an independent and *controlled* variable, and $Consent_I$ (consent for the use of the inference) as a dependent variable. We repeatedly measure users' consent for the use of one inference of each category: 1) arts, 2) ads, and 3) protected. Experiment 2 is between-subjects with inference *Category* as an independent and *controlled* variable and $Consent_A$ (consent for the use of their answers after inference awareness) as a dependent variable. In both experiments, we administered the post-experiment survey, where we measured participants' *Sentiment* of the system's data practices and their feeling of being *Informedness* when providing consent to use their answer. *Correctness* of inference is a nested and unbalanced factor in both experiments that we cannot experimentally control for but account for its imbalance in our analysis. Table 4 summarizes the independent and dependent variables in both experiments.

3.8 Ethical Considerations

Understanding privacy preferences with AI inferences requires us to collect data and make participants aware of inferences of varying sensitivity (ranging from least sensitive, such as their hobbies, to very sensitive, such as their employment status). To reduce discomfort and minimize any harm to participants, we prevented PAR from asking the most sensitive questions from CPS (e.g., questions about disability status and questions closely related to socio-economic status, such as income). We identified these questions based on the iterative annotation and coding process described in Section 3.1.4. People's levels of comfort can be different and nuanced, and some included questions could still be uncomfortable. We acknowledged this in our study consent form and gave participants the option to leave the experiments at any time should they decide so.

Acknowledging the sensitivity of the data about participants from the experiments, we stored all data on secure institution's

Table 4: Overview of experimental design for the two experiments and the main dependent and independent variables in them.

	Variable	Role	Type	Levels
Experiment 1 (RQ 1,4)	Inference category (balanced)	independent	Categorical	3
	Inference correctness (unbalanced)	independent	Categorical	2
	Consent to use *inferred attribute*	dependent	Categorical	4
	Consent to use *inferred arts attribute*	dependent	Categorical	4
Experiment 2 (RQ 2,3,4)	Inference category (balanced) (3 inference, 1 no-inference)	independent	Categorical	4
	Inference correct (unbalanced)	independent	Categorical	2
	Consent to use the most informative *answer* of the inference	dependent	Categorical	4
	Sentiment	dependent	Categorical	5 (Likert)
	Informedness	dependent	Categorical	5 (Likert)
	Consent to use *explicitly* answered arts attribute	dependent	Categorical	4

servers, and the storage protocols were reviewed by our institution's review board (IRB). While we collect sensitive participant data, our questions are adopted from the US Census, which already anonymizes participant responses up to 100,000 responses in a statistical area. As such, our data and inferences are at least as anonymous as the US Census responses. Our study was deemed exempt from ongoing IRB oversight.

3.9 Limitations

Our method comes with limitations, most of which are due to using a hypothetical public art recommendation system to conduct lab user studies. For example, PAR focuses on public arts recommendations, but privacy preferences can vary considerably by context [59] (e.g., health, social media, online dating), the quality of online service, and the actual benefits of personalization [1]. PAR does not provide users with actual benefits of personalization due to the short interaction and the focus on inferences. While our focus on the "cold-start" problem precludes us from studying the benefits of personalization and inferences based on data collected through long-term interaction, it is a reasonable first step towards studying consent for using inferences. Inferences generated from long-term behavior activity can compound studying privacy decision-making due to the difficulty of linking behavior traces with inferences.

Next, our method leaves the possibility for participants to provide false information to our recommendation system. Although this does not affect the order of questions that PAR asks, it could affect the accuracy of our inferences. Thus, we accounted for the correctness of inferences, confirming them again with participants. In our method, we communicated to participants that their data is strictly confidential and only accessible to research team members. During the debriefing stage, we reminded participants about the value of the research of knowing if the inference was correct.

Despite confirming with participants about the accuracy of their data, some participants could still intentionally provide wrong answers to the system. Users can provide wrong answers to online services if they feel compelled to provide information but do not see benefits [74] or to safeguard sensitive data [90]. If participants intentionally provided wrong answers in our study, their consent decisions would also be in the context of wrong answers, and we can not know how they may have provided consent for the use of correct data. Future work can consider strategies for establishing correctness, such as triangulation with secondary data sources (e.g., additional surveys or platform collected data) or predicting the likelihood of the user's answer being true [87] to decide whether the information can be used for personalization.

We generated inferences, made participants aware of such inferences, and asked for their consent to use their *inference* or their *most informative answer for the inference* (Section 3.2). Asking participants to provide consent to use their answers after awareness of the inference could be a form of digital resignation for participants who are not comfortable with even the generation of inferences. Such digital resignation relates to the discomfort experienced by users when compelled to disclose personal data for using online services [96].

We also note that there could be a gap between the accuracy of our system and those of major tech companies, which have access to abundant user data. However, during our pilot tests of the study software, participants did not make explicit complaints about the accuracy, and prior work suggests that inference about users is difficult even for platforms with extensive user data [114]. Moreover, we evaluated our algorithm offline on the CPS dataset, and the average accuracy of inferences for various user attributes was 65% across inferences on all personal attributes, with the highest accuracy of 100% across some personal attributes.

4 RESULTS

Here, we present the results from the two experiments. We first validated that participants were willing to answer questions that our questions selection algorithm asked. Figure 5 shows the questions that PAR asked participants in both experiments. Two of the most frequently asked questions were about the participant's age and state of residence, as they are highly predictive of several other attributes of participants. The participants answered 99.9% of all questions that PAR asked them across the two experiments. Furthermore, the system had an overall average accuracy of 48% correct inferences across both experiments.

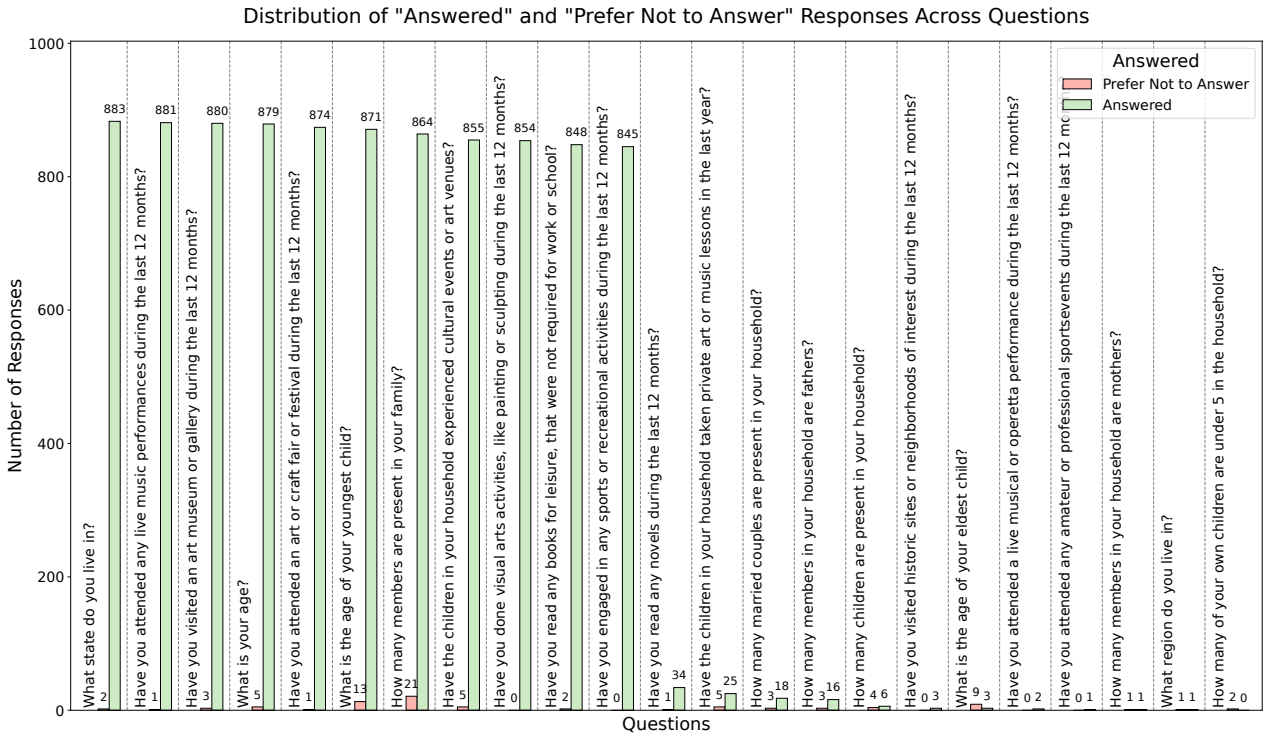


Figure 5: Bar plot illustration of all questions that the participants answered and chose not to answer.

4.1 Experiment 1: Privacy Preferences for Inferred Personal Attributes

In Experiment 1, PAR asked each participant to provide their consent for the use of one inference from each of the three categories. The contingency table (Table 5) details the breakdown of correct and incorrect inferences by inference category across the 333 participants who took part in this experiment. Figure 6 shows the frequency of all the attributes that the system inferred about participants and the proportion of correct inferences for each attribute.

Table 5: Contingency table across inference category and inference correctness for Experiment 1.

Inference category	Inference correct	Inference incorrect	Total	Inference accuracy
Category arts	213	120	333	63%
Category ads	160	173	333	45%
Category protected	130	203	333	39%

4.1.1 Consent decisions for inferred attributes. We found a statistically significant effect of *Category* on $Consent_I$ in our analysis ($F(2, 617) = 6.92; p < 0.01$) after controlling for the effect of *Correctness* across participants. Mean consent level when *Category* was *arts* (mean = 1.22) and *ads* (mean = 1.22) was higher than for *protected* category (mean = 1.0) ($F(2, 631) = 29.5; p < 0.001$). We found a medium effect size for the difference between the categories *arts* and *protected* (Cohen's $d = 0.675$) and a small effect size for the difference between the categories *ads* and *protected* (Cohen's $d = 0.463$). Thus, participants provided on an average *higher consent level* for the use of their inferred attributes from the arts and ads categories than attributes from the protected category. Figure 7a shows the distribution of inferences across the three categories.

We also found a statistically significant effect of *Correctness* on $Consent_I$ ($F(1, 332) = 40; p < 0.001$). Correct inferences had a higher consent level (mean = 1.50, SD = 1.01) than incorrect ones (mean = 0.83, SD = 1.07, Cohen's $d = 0.593$). Thus, participants were more likely to provide higher consent levels (broader) for the use of their inferred attributes if those inferences were correct than if they were wrong. Table 6 summarizes the coefficients for the ANOVA test in our analysis.

Figure 7 shows the distribution of consent decisions for inferred attributes across arts, ads, and protected categories (a) and by correctness (b). In Figure 7a, the box ranges indicate similar 25th and 75th percentile for all categories. The distribution of the dots shows that protected inferences have more participants choosing the lowest consent level ("do not use and store at all") compared to other conditions. Variation of consent by inference category and correctness (Figure 7b) reveals more clear differences. When inferences are correct, the boxes are relatively higher than for incorrect inferences, indicating that more participants are choosing higher levels.

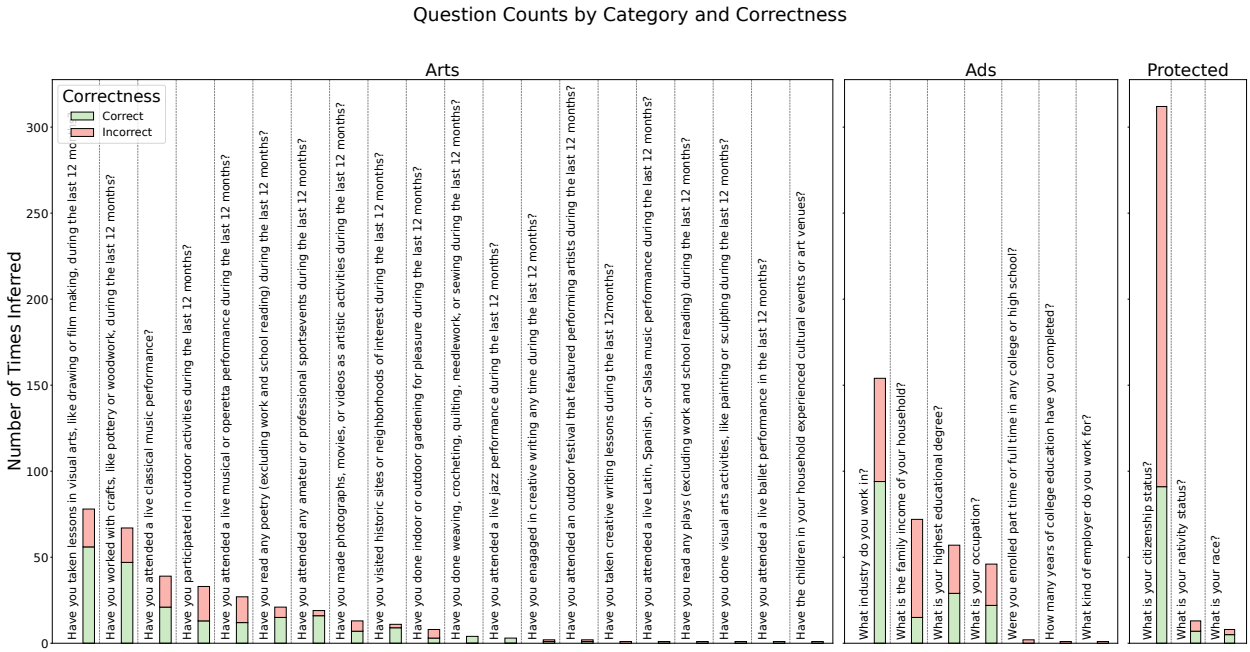


Figure 6: Stacked bar-plot distribution of user attributes that the system inferred about the participants in Experiment 1. Green bars represent the number of times PAR inferred that attribute correctly across participants. Pink bars represent the number of times PAR inferred that attribute incorrectly across participants.

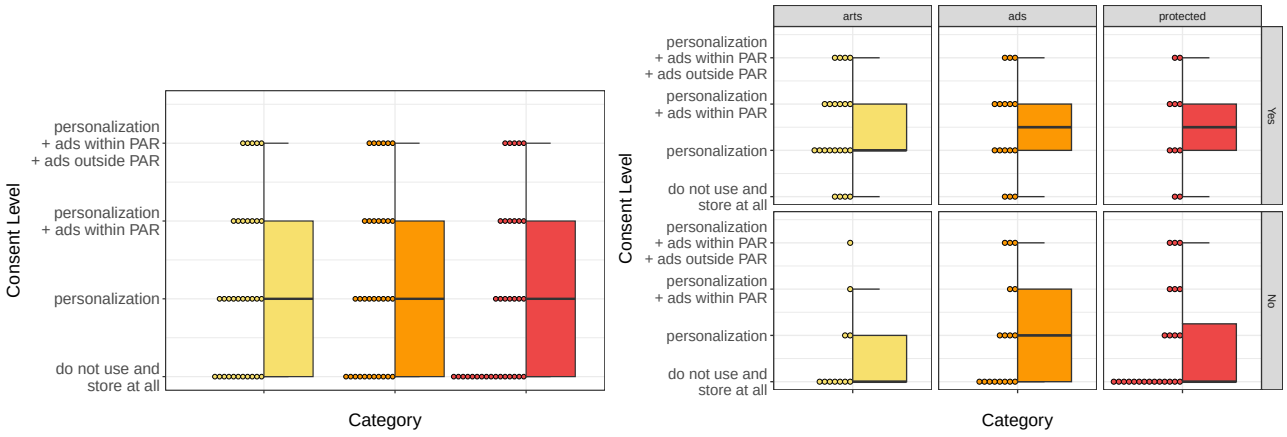


Figure 7: Variation of consent to use inferred attributes against inference category (a) and by inference category and correctness (b). Dots show the number of participants that indicated the corresponding consent level. Each dot represents 10 participants. Bounds of boxes indicate the 25th and 75th quartile of consent levels, and the middle line indicates the median (middle).

4.1.2 Privacy concerns. From the answers to the post-experiment survey asking them about their concerns (1 - Extremely concerned, 5 - Not concerned at all), participants were moderately concerned about the system invading (*Invasive*) their privacy (mean = 2.36, SD=1.02). Participants were also moderately concerned (1 - Extremely concerned, 5 - Not concerned at all) about the misuse (*Misuse*) of their private information by the system (mean=2.22, SD=1.1). Participants were generally comfortable (*Comfort*) (1 - Extremely comfortable, 5 - Not comfortable at all) with the system learning information about them (mean=1.91, SD=0.92).

Table 6: Coefficients for ANOVA for effect of *Category* on $Consent_I$ for inferred attributes.

Condition tests				
Variable	F	Df	Df.res	Pr(>F)
Inference category	29.56	2	631.64	5.34e-13
Correctness	40.83	1	332.57	5.58e-10
Inference category:Correctness	0.16	2	530.83	0.85
Contrast tests				
contrast	estimate	SE	Df	p-value
$Category_{arts}$ - $Category_{ads}$	37.18	14.98	633.76	1.33e-02
$Category_{arts}$ - $Category_{protected}$	118.58	15.72	630.09	4.77e-13
$Category_{ads}$ - $Category_{protected}$	81.40	15.37	630.72	3.299340e-07

4.2 Experiment 2: Privacy Preferences for Explicitly Provided Personal Attributes Used in Inferences

In this experiment, participants provided consent to use their answers after becoming aware of an inference from one of the categories or provided consent to use their answer without awareness of any inference (when *Category* is *none*). The contingency table (Table 7) details the breakdown of the observations across the independent variables *Category* and *Correctness*. Figure 8 shows the frequency of all the attributes that the system inferred about participants and the proportion of correct inferences for each attribute.

Table 7: Contingency table across inference category and inference correctness for Experiment 2

Category	Inference correct	Inference incorrect	Total	Inference accuracy
Category none	-	-	136	-
Category arts	98	38	136	72%
Category ads	61	75	136	52%
Category protected	55	81	136	40%

4.2.1 Variation of consent decisions for the use of answered attributes with awareness of inference across Category. $Consent_A$ measured participants' consent to use the *most informative answer for the inference* across *Category* of inference that PAR made participants aware of. We found a statistically significant effect of *Category* of inference awareness on participant's consent decision for use of their *most informative answer for the inference* ($F(3, 538) = 2.74; p < 0.05$). Our post-hoc tests found a statistically significant difference in participants' consent levels between $Category = none$ and $Category = protected$ ($estimate = 49.42, SE = 18.41, p < 0.05$). We found a small effect size for this difference (Cohen's $d = 0.3262$).

4.2.2 Variation of consent decisions for use of answered attributes across inference Category (arts, ads, protected) and Correctness. In conditions where participants were made aware of inferences, we compared the effect of $Category \times Correctness$ on $Consent_A$. Our two-way mixed ANOVA model ($Category \times correctness$) found a statistically significant interaction effect of *Category* and *Correctness*. ($F(2, 402) = 3.031; p < 0.05$). Our post-hoc tests found a statistically significant difference in participants' consent levels between $Category = arts$, $Correct = yes$ and $Category = protected$, $Correct = no$ ($estimate = 44.51, SE = 15.75, p < 0.1$). We found a small effect size (Cohen's $d = 0.16$) for this contrast.

Table 8: Coefficients for one-way ANOVA for effect of transparency of inference *Category* on $Consent_A$. Table only shows statistically significant contrasts in the post-hoc tests

Condition tests				
Variable	F	Df	Df.res	Pr(>F)
Category	2.74	3	538	0.042842
Contrast tests				
contrast	estimate	SE	Df	p-value
$Category_{none}$: $Category_{protected}$	49.42	18.41	538	0.045

Table 9: Coefficients for two-way ANOVA for effect of transparency of inference *Category* and *Correctness* on $Consent_A$ across inference categories. Table only shows statistically significant contrasts in the post-hoc tests

Condition tests				
Variable	F	Df	Df.res	Pr(>F)
Category	1.57	2	402	0.21
Correctness	0.62	1	402	0.043
Category:Correct	3.03	2	402	0.049
Contrast tests				
contrast	estimate	SE	Df	p-value
$Arts, Correct$: $Protected, Incorrect$	44.51	15.75	402	0.074

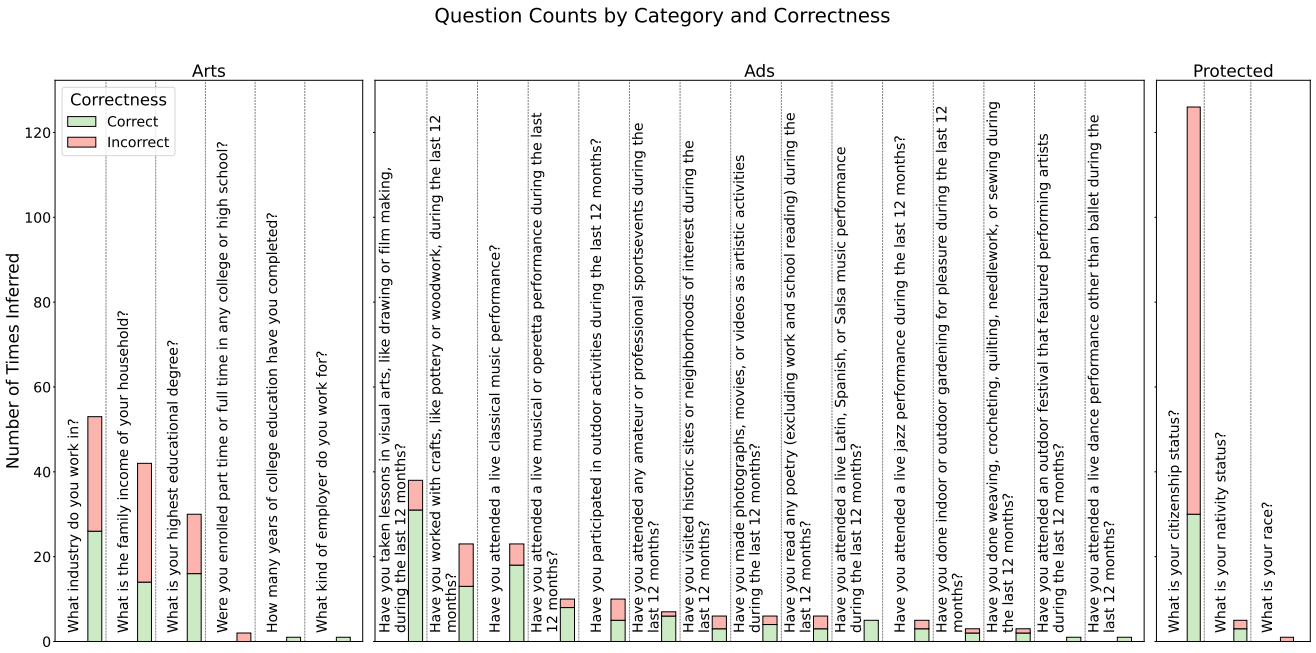


Figure 8: Figure shows the bar-plot distribution of user attributes that the system inferred about the participants in Experiment 2. Each bar represents the number of times an attribute was inferred.

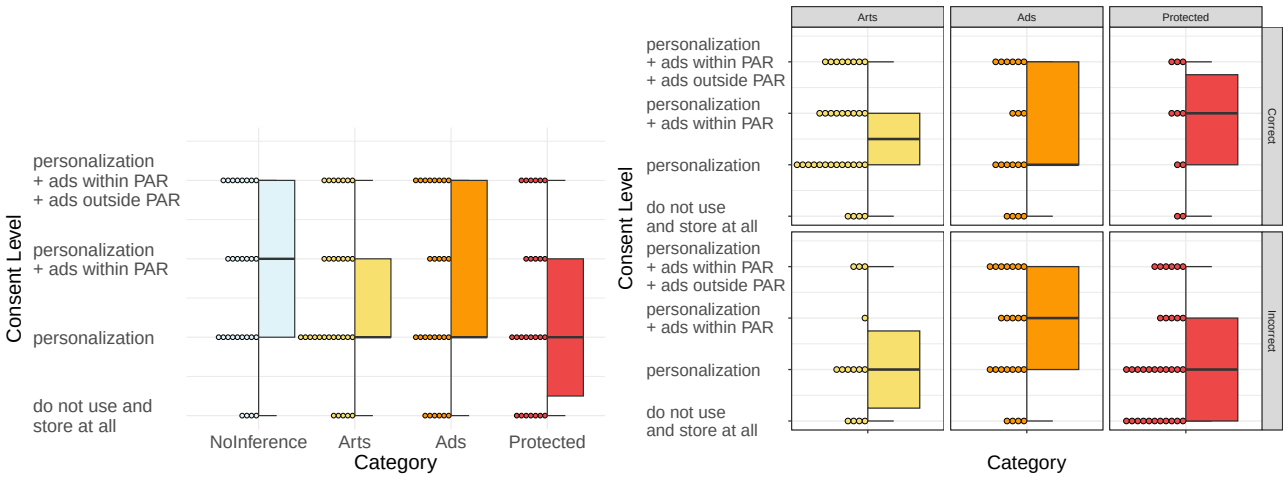


Figure 9: Variation of consent to use inferred attributes against inference category (a) and by inference category and correctness (b). Dots indicate the number of participants who chose the corresponding consent level. Each dot represents 5 participants. Bounds of boxes indicate the 25th and 75th quartile of consent levels, and the middle line indicates the median (middle).

Figure 9 shows consent decisions by *Category* alone (left) and by *Category* and *Correctness* across inference categories. From Figure 9a, we observe that the quartiles for the *none*, *arts*, and *ads* categories are comparable. However, the quartiles for the *protected* category are one level lower. The dots against the boxes indicate the same trend. Figure 9b shows the variation of consent decisions by condition and correctness of inferences in the three inference conditions, suggesting a trend similar to Experiment 1. Correct inferences have higher quartiles for consent decisions than incorrect inferences, and this difference is most pronounced for protected inferences. However, surprisingly, we see higher quartiles (consent levels) for the ads categories from both figures.

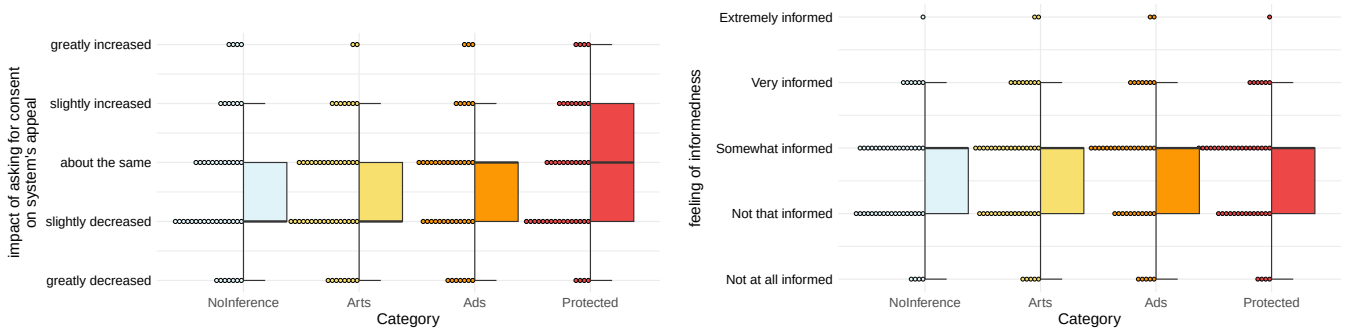


Figure 10: Figure shows participants' appeal (sentiment) of the system's data practices when it asked for consent to use their answer after showing an inference (a) and participants' overall feeling of informedness if they provided consent to use their answer after seeing an inference (b) after providing consent to use the *most informative answer for the inference*. Each dot represents 5 participants.

For *H4*, we find that the mean consent level for explicitly provided arts attributes (mean = 1.74) is higher than the mean consent level for $inferred$ arts attributes (mean = 1.43), and this difference is statistically *significant* (95% CI $[0.006, 0.612]$ $t(160) = 2.01; p < 0.05$) (Figure 11). We found a small effect size (Cohen's $d = 0.3$) for this difference. Thus, we find support for hypothesis (H4) that consent levels for the same type of attributes are *lower* when they are inferred than when they are explicitly provided by participants.

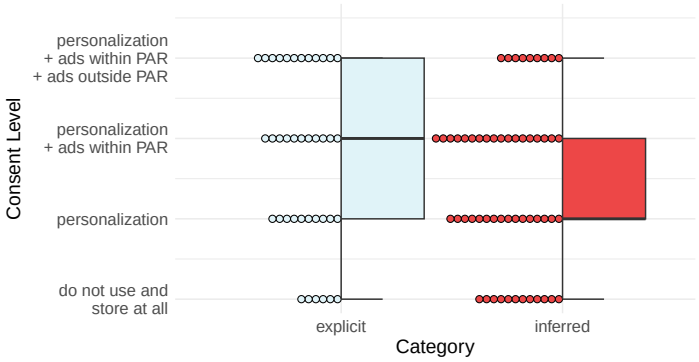


Figure 11: Half-dot half-box plot showing consent levels for *explicitly answered* vs inferred *arts* attributes from Experiment 1 and 2. The blue box and its median line showed slightly higher quartile ranges (more proportion of participants in higher consent levels) when PAR asked for consent to use answered arts attributes compared to when they are inferred by PAR (red). Each dot represents 3 participants.

4.2.3 Privacy concerns and user sentiment. From the answers to the post-experiment survey, participants were moderately concerned about the system invading their privacy (*Invasive*) (mean = 2.36, SD=1.02) (1 - Extremely concerned, 5 - Not concerned at all). Participants were also moderately concerned about misuse (*Misuse*) of their private information (mean = 2.22, SD=1.0) (1 - Extremely concerned, 5 - Not concerned at all) by the system. Participants were generally comfortable with the system learning information about them (mean = 1.9, SD=0.92) (1 - Extremely comfortable, 5 - Not comfortable at all). We observe that the ratings are similar to the previous experiment, showing a moderate privacy concern.

For participants' perceived $Informedness$ (1 - not at all informed, 5 - extremely informed) regarding the system's data practices when PAR asked their consent to use their answer, their mean ratings did not exhibit any trend: no-inference (mean = 2.62, SD = 0.9), arts (mean = 2.74, SD = 1.01), ads (mean = 2.79, SD = 1.02), protected (mean = 2.69, SD = 0.87). For participants' sentiment (*sentiment*) towards the system's data practices when PAR asked their consent to use their answer across inference *Category* (1 - greatly decreased, 5 - greatly increased), we found the mean rating to be highest for the protected condition (mean = 2.78, SD = 1.11) and lowest for the arts (mean = 2.56, SD = 1.07) with no-inference (mean = 2.61, SD = 1.13) and ads (mean = 2.65, SD = 1.13) in between. However, we did not find a statistically significant effect for any of these differences in our analysis (H3).

4.3 Summary of Results from Experiments 1 & 2

We find that participants chose lower consent levels for the use of their most sensitive (protected) and incorrectly inferred attributes (H1.1 and H1.2) (from Experiment 1). Participants who were made aware of an inference from their explicitly provided answer also chose lower consent levels for the use of such answers than participants who chose consent without the awareness of any inference(H2.1). Further, participants who became aware of protected and incorrect inferences were most likely to choose lower consent levels for their answers compared to participants who were made aware of other types of inferences (H2.2). We did not observe any effect of the awareness of different inference categories on participants' feeling of being informed ($Informedness$) about the system's data practices. We found a marginal but statistically non-significant increase in participants' *sentiment* towards the system's data practices after awareness of protected inferences compared to other categories. Finally, we found that participants were more likely to

give higher consent decisions for the use of their explicitly provided arts attribute answers than when PAR inferred the arts attributes about them (H4). This provides preliminary evidence for differential treatment for consent to use inferences vs explicitly provided data.

5 DISCUSSION

Although users are generally aware that platforms learn from their behaviors to infer their interests and demographics, our findings generate new knowledge about users' privacy preferences for their inferred data in the context of personalization [66, 85]. Our findings show how awareness of AI inferences in the context of users' provided data impacts their privacy preferences.

5.1 Differential Consent Levels for Explicitly Provided and Inferred Attributes by Sensitivity and Accuracy

Awareness of *protected* (most sensitive) and incorrect inferences made participants choose lower consent decisions for the use of such inferences and lower consent decisions for the use of their explicitly provided answers. Users could be familiar with inferences of less sensitive attributes (*arts* and *ads*) through past experiences with personalized services and targeted ads [83]. While services use inferred protected attributes for algorithmic decision-making [110], it is less intuitive for users to link their provided data or the targeted ads they see to inferences on *protected* attributes [113]. This could have surprised participants once they learned about the inferences and impacted their consent decisions.

Users provided lower consent decisions for the use of inferred arts attributes compared to explicitly provided arts attributes, suggesting a differential treatment towards consent for the use of inferred data compared to explicitly provided data. Different consent decisions for protected attributes and when the same type of attributes are inferred versus explicitly provided could suggest a lack of user knowledge regarding inferences. Efforts to enhance people's inference literacy, a term that Warshaw et al. [113] coined to refer to digital literacy [81] on how companies use and store inferences could help users make informed choices for the use of their data in the context of AI inferences. We need to rethink existing consent mechanisms (such as notice and choice) [25] as they do not provide scope for raising people's digital literacy.

Wrong inferences made users choose lower consent levels both for such inferences and for their explicitly provided answers after awareness of such inferences. One reason could be that participants wanted to minimize any unwanted content, ads, or risks, such as breach of false personal information. Another possible explanation is the mismatch between how users perceive the use of their data and advertisers' actual data usage. Advertisers often label users for their potential future behaviors [85], which could be correct or incorrect. However, users are unlikely to be aware that inferences could be wrong but still useful to advertisers [109].

We also observed that consent decisions were highly variable across individuals. ANOVA for consent decisions had a relatively low F-value, indicating high in-group variance. This suggests that other factors like demographics [122], context of use [57], and prior experiences of privacy violations [69] are likely contributors to consent decisions just like they affect consent for explicitly provided data [16]. For example, users who faced adverse consequences of privacy violations could choose high consent levels due to resignation towards existing data practices [113] instead of enthusiastically wanting inferences for personalization. Thus, understanding the privacy needs of different user groups will be important in designing more inclusive consent interfaces for the use of inferences.

We also note that although it is possible that participants could have provided false data in our study, such behavior could be part of realistic interactions when users may have heightened privacy concerns but still need to provide data [17, 74, 119]. Thus, false data forms a part of our study because participants evaluate inferences with awareness of the accuracy of the data that they provide. However, intentional false data could be a symptom of consent practices that do not adequately satisfy user needs. Awareness of inferences, especially wrong inferences, can make users more protective of their data, which could, in turn, impact the extent to which users provide false data to services.

5.2 Designing Consent Mechanisms for Inferred User Attributes for Personalization

Scholarship and audits on privacy concerns of user attributes collected with their awareness have informed consent mechanisms for such attributes [56, 83, 91], but due to the black-box nature of inferences and limited studies about them [84, 121], it is very challenging to expand consent mechanisms for inferred personal attributes [10, 85, 115]. Our experiments provide initial directions for expanding the discourse on consent to the use of inferred personal attributes by showing that users indeed provide different consent levels for inferences in different categories (arts, ads, and protected) and for correct and incorrect inferences. Furthermore, participants who learned about inferences also gave restrictive privacy decisions for their explicitly provided answers compared to participants who did not learn about any inferences. This is in line with prior studies where users' knowledge of inferences from their data makes them more aware of their privacy preferences and online behaviors [9, 116]. Consent mechanisms need to accurately convey to users how one's data can be used to infer new information about them and ask for their consent—instead of passively collecting data and using them for inference.

However, as the discourse around notice and consent has shown, many consent popups and data privacy settings are currently designed in a way that is burdensome and useless for users even for existing data [12, 92]. Privacy scholars have pointed out that many choices provided by companies are unusable [93] and prone to mislead users through misguided messaging or designs ("dark patterns") [68, 98]. The design space to improve notice and choice consent along the dimensions of timing, channel, modality, and control in online interactions [93, 94] suggests opportunities for building consent with AI inferences. AI inferences may likely require additional design considerations, such as appropriate communication strategies to convey inferences.

Data aggregation across platforms, coupled with the increasing accuracy of AI algorithms, dramatically increases the number of personal attributes that online services can infer about their users and the need for consent to use such attributes. Simply adding every possible inference to existing notice and choice mechanisms could

lead to cognitive burden and make consent even less informed [71]. More innovative methods beyond notice and choice may be needed to enable informed consent of users. For example, the improved conversational capabilities of chatbots could be leveraged for more interactive privacy awareness [41] than simply burdening users with a long list of notices about their privacy. Similarily, better regulations around "privacy-protective designs and practices" can contribute towards consent-friendly use of AI-powered applications [112]. Regulations can also lead to mechanisms that educate the public about AI so that they can make more informed decisions about their privacy [49].

Our findings also suggest that transparency of inferences and building consent mechanisms around it could also benefit companies and not just users. When asked about the system's appeal in Experiment 2, to our surprise, we found that transparency around protected inferences evoked the highest appeal for the system's data practices across all conditions. A possible explanation is that while users are already aware of the existence of inferences [85, 121], making a sensitive inference transparent and then asking users for consent about their data positively impacted their impression of the system. This hints at the possibility that consent mechanisms that are grounded in more meaningful and enthusiastic consent [58] can contribute to enhanced trust between users and companies.

6 CONCLUSION AND FUTURE WORK

We study the impact of AI inferences on users' consent decisions for the system's usage of their information. We conducted two experiments where participants (N=877) first answered questions about themselves for personalized public arts recommendations. Then, participants indicated their level of consent to let the system use their inferred information or explicitly provided answers to generate inferences. We found that they provide more restrictive consent decisions when they learn that highly sensitive inferences can be generated from their answers. This consent behavior holds true both for inferred user attributes and their explicitly provided answers after they become aware of inferences.

Our findings show that given a choice, users would choose their privacy preferences for inferences similar to explicit data instead of an all acceptance or rejection of inferences. This suggests a positive attitude towards personalization, but it requires online services to rethink their data practices and privacy regulations so that they center users' current level of understanding and privacy needs regarding inferences.

We studied consent to use inferences in a limited setting of three inferences per user and consent to use one answer in the presence of one inference. An important research direction is designing practical consent mechanisms that make it easier for users to mark their preferences for both their data and inferences made about them for many inferences and in the presence of implicit data. Also, we studied inferences in the context of a public arts recommender system, which reflects a scenario of using common entertainment applications. Future work should look into other contexts, such as how social media's inferences impact users' perceptions of the platform's use or how users prefer to give consent to the use of inferences made for health reasons as consent to information varies greatly by context [59].

Our work also opens up directions to study consent around the storage of inferences and its potential use by algorithms, which we did not cover in our research but is an increasingly important topic due to algorithmic decision-making [10]. Lastly, due to the misconceptions around AI inferences [85, 113] users could be thought to be taking actions incommensurate with their privacy preferences, but a careful education of users about AI can enable them to align their online behaviors with their privacy preferences.

ACKNOWLEDGMENTS

We thank our pilot participants for helping us evaluate the study interface. We also thank the anonymous participants on Prolific who participated in our study, without whom our research would not have been possible. We thank Divya Ramesh for the initial discussions about the idea. We also thank Snehal Prabhudesai for their feedback on project discussions and data processing. We also thank Yasha Iravantchi, Rebecca Krosnick, and Anjali Singh for their feedback on the manuscript. The work was funded by the Department of Energy (DoE) grant number DE-SC0021398, and Jane Im was additionally supported by the Barbour Scholarship.

REFERENCES

[1] Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. 2015. Privacy and human behavior in the age of information. *Science* 347, 6221 (2015), 509–514. https://doi.org/10.1126/science.aaa1465

[2] Alessandro Acquisti and Jens Grossklags. 2005. Privacy and rationality in individual decision making. *IEEE security & privacy* 3, 1 (2005), 26–33. https://doi.org/10.1109/MSP.2005.22

[3] Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. 2019. Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Biased Outcomes. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 199 (nov 2019), 30 pages. https://doi.org/10.1145/3359301

[4] McKane Andrus, Elena Spitzer, Jeffrey Brown, and Alice Xiang. 2021. What We Can't Measure, We Can't Understand: Challenges to Demographic Data Procurement in the Pursuit of Fairness. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) *(FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 249–260. https://doi.org/10.1145/3442188.3445888

[5] McKane Andrus, Elena Spitzer, Jeffrey Brown, and Alice Xiang. 2021. What We Can't Measure, We Can't Understand: Challenges to Demographic Data Procurement in the Pursuit of Fairness. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) *(FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 249–260. https://doi.org/10.1145/3442188.3445888

[6] Ankur Ankan and Abinash Panda. 2015. pgmpy: Probabilistic graphical models using python.

[7] Nikola Banovic and John Krumm. 2018. Warming Up to Cold Start Personalization. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 124 (jan 2018), 13 pages. https://doi.org/10.1145/3161175

[8] Gaurav Bansal, Fatemeh "Mariam" Zahedi, and David Gefen. 2010. The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. *Decision Support Systems* 49, 2 (2010), 138–150. https://doi.org/10.1016/j.dss.2010.01.010

[9] Natã M. Barbosa, Zhuohao Zhang, and Yang Wang. 2020. Do Privacy and Security Matter to Everyone? Quantifying and Clustering User-Centric Considerations about Smart Home Device Adoption. In *Proceedings of the Sixteenth USENIX Conference on Usable Privacy and Security (SOUPS'20)*. USENIX Association, USA, Article 22, 19 pages.

[10] Miranda Bogen, Aaron Rieke, and Shazeda Ahmed. 2020. Awareness in Practice: Tensions in Access to Sensitive Attribute Data for Antidiscrimination. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Barcelona, Spain) *(FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 492–500. https://doi.org/10.1145/3351095.3372877

[11] Rico Bornschein, Lennard Schmidt, and Erik Maier. 2020. The Effect of Consumers' Perceived Power and Risk in Digital Information Privacy: The Example of Cookie Notices. *Journal of Public Policy & Marketing* 39, 2 (2020), 135–154. https://doi.org/10.1177/0743915620902143 arXiv:https://doi.org/10.1177/0743915620902143

[12] Elijah Robert Bouma-Sims, Megan Li, Yanzi Lin, Adia Sakura-Lemessy, Alexandra Nisenoff, Ellie Young, Eleanor Birrell, Lorrie Faith Cranor, and Hana Habib. 2023. A US-UK Usability Evaluation of Consent Management Platform Cookie Consent Interface Design on Desktop and Mobile. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 163, 36 pages. https://doi.org/10.1145/3544548.3580725

[13] Fred H Cate. 2016. The failure of fair information practice principles. In *Consumer Protection in the Age of the 'Information Economy'*. Routledge, Chapter 13, 341–377.

[14] Farah Chanchary and Sonia Chiasson. 2015. User Perceptions of Sharing, Advertising, and Tracking. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*. USENIX Association, Ottawa, 53–67. https://www.usenix.org/conference/soups2015/proceedings/presentation/chanchary

[15] Tsai-Wei Chen and S. Shyam Sundar. 2018. This App Would Like to Use Your Current Location to Better Serve You: Importance of User Assent and System Transparency in Personalized Mobile Services. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174111

[16] Hichang Cho, Jae-Shin Lee, and Siyoung Chung. 2010. Optimistic bias about online privacy risks: Testing the moderating effects of perceived controllability and prior experience. *Computers in Human Behavior* 26, 5 (2010), 987–995. https://doi.org/10.1016/j.chb.2010.02.012 Advancing Educational Research on Computer-supported Collaborative Learning (CSCL) through the use of gStudy CSCL Tools.

[17] Julien Cloarec. 2022. Privacy controls as an information source to reduce data poisoning in artificial intelligence-powered personalization. *Journal of Business Research* 152 (2022), 144–153.

[18] Amit Datta, Michael Carl Tschantz, and Anupam Datta. 2015. Automated Experiments on Ad Privacy Settings. *Proc. Priv. Enhancing Technol.* 2015, 1 (2015), 92–112. https://doi.org/10.1515/popets-2015-0007

[19] Pam Dixon and Robert Gellman. 2014. The scoring of America.

[20] Lisa A. Elkin, Matthew Kay, James J. Higgins, and Jacob O. Wobbrock. 2021. An Aligned Rank Transform Procedure for Multifactor Contrast Tests. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 754–768. https://doi.org/10.1145/3472749.3474784

[21] K Anders Ericsson. 2017. Protocol analysis. *A companion to cognitive science* (2017), 425–432.

[22] Motahhare Eslami, Sneha R. Krishna Kumaran, Christian Sandvig, and Karrie Karahalios. 2018. Communicating Algorithmic Process in Online Behavioral Advertising. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174006

[23] Florian M. Farke, David G. Balash, Maximilian Golla, Markus Dürmuth, and Adam J. Aviv. 2021. Are Privacy Dashboards Good for End Users? Evaluating User Perceptions and Reactions to Google's My Activity. In *30th USENIX Security Symposium (USENIX Security 21)*. USENIX Association, 483–500. https://www.usenix.org/conference/usenixsecurity21/presentation/farke

[24] Rachel F Fefer. 2019. Data flows, online privacy, and trade policy.

[25] Yuanyuan Feng, Yaxing Yao, and Norman Sadeh. 2021. A Design Space for Privacy Choices: Towards Meaningful Privacy Control in the Internet of Things. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 64, 16 pages. https://doi.org/10.1145/3411764.3445148

[26] Casey Fiesler, Cliff Lampe, and Amy S. Bruckman. 2016. Reality and Perception of Copyright Terms of Service for Online Content Creation. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (San Francisco, California, USA) *(CSCW '16)*. Association for Computing Machinery, New York, NY, USA, 1450–1461. https://doi.org/10.1145/2818048.2819931

[27] Sarah Flood, Miriam King, Renae Rodgers, Steven Ruggles, J. Robert Warren, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Megan Schouweiler, and Michael Westberry. 2023. IPUMS CPS: Version 11.0. [dataset]. https://doi.org/10.18128/D030.V11.0

[28] Asbjørn Følstad, Cecilie Bertinussen Nordheim, and Cato Alexander Bjørkli. 2018. What Makes Users Trust a Chatbot for Customer Service? An Exploratory Interview Study. In *Internet Science*, Svetlana S. Bodrunova (Ed.). Springer International Publishing, Cham, 194–208.

[29] Hamish Fraser, Enrico Coiera, and David Wong. 2018. Safety of patient-facing digital symptom checkers. *The Lancet* 392, 10161 (2018), 2263–2264.

[30] Satish Garla, Albert Hopping, Rick Monaco, and Sarah Rittman. 2013. What do your consumer habits say about your health? Using third-party data to predict individual health risk and costs.

[31] California Attorney General. 2020. Ofce of the California Attorney General. 2020. California Consumer Privacy Act (CCPA): Final Text of Proposed Regulations. https://oag.ca.gov/sites/all/files/agweb/pdfs/privacy/oal-sub-final-text-of-regs.pdf

[32] Nina Gerber, Paul Gerber, and Melanie Volkamer. 2018. Explaining the privacy paradox: A systematic review of literature investigating privacy attitude and behavior. *Computers & security* 77 (2018), 226–261.

[33] Phillipa Gill, Vijay Erramilli, Augustin Chaintreau, Balachander Krishnamurthy, Konstantina Papagiannaki, and Pablo Rodriguez. 2013. Best Paper – Follow the Money: Understanding Economics of Online Aggregation and Advertising. In *Proceedings of the 2013 Conference on Internet Measurement Conference* (Barcelona, Spain) *(IMC '13)*. Association for Computing Machinery, New York, NY, USA, 141–148. https://doi.org/10.1145/2504730.2504768

[34] Jyotirmoy Gope and Sanjay Kumar Jain. 2017. A survey on solving cold start problem in recommender systems. In *2017 International Conference on Computing, Communication and Automation (ICCCA)*. IEEE, 133–138. https://doi.org/10.1109/CCAA.2017.8229786

[35] David A Grant. 1948. The latin square principle in the design and analysis of psychological experiments. *Psychological bulletin* 45, 5 (1948), 427.

[36] Colin M. Gray, Cristiana Santos, Nataliia Bielova, Michael Toth, and Damian Clifford. 2021. Dark Patterns and the Legal Requirements of Consent Banners: An Interaction Criticism Perspective. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 172, 18 pages. https://doi.org/10.1145/3411764.3445779

[37] Dorota Głowacka and Karolina Iwańska. 2021. Algorithms of trauma: new case study shows that Facebook doesn't give users real control over disturbing surveillance ads. Retrieved December 9, 2022 from https://en.panoptykon.org/algorithms-of-trauma

[38] Hana Habib, Megan Li, Ellie Young, and Lorrie Cranor. 2022. "Okay, Whatever": An Evaluation of Cookie Consent Interfaces. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 621, 27 pages. https://doi.org/10.1145/3491102.3501985

[39] Hana Habib, Yixin Zou, Aditi Jannu, Neha Sridhar, Chelse Swoopes, Alessandro Acquisti, Lorrie Faith Cranor, Norman Sadeh, and Florian Schaub. 2019. An Empirical Analysis of Data Deletion and Opt-Out Choices on 150 Websites. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*. USENIX Association, Santa Clara, CA, 387–406. https://www.usenix.org/conference/soups2019/presentation/habib

[40] Moritz Hardt, Eric Price, Eric Price, and Nati Srebro. 2016. Equality of Opportunity in Supervised Learning. In *Advances in Neural Information Processing Systems*, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (Eds.), Vol. 29. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2016/file/9d2682367c3935defcb1f9e247a97c0d-Paper.pdf

[41] Hamza Harkous, Kassem Fawaz, Kang G. Shin, and Karl Aberer. 2016. PriBots: Conversational Privacy with Chatbots. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*. USENIX Association, Denver, CO. https://www.usenix.org/conference/soups2016/workshop-program/wfpn/presentation/harkous

[42] David G. Herr. 1986. On the History of ANOVA in Unbalanced, Factorial Designs: The First 30 Years. *The American Statistician* 40, 4 (1986), 265–270. http://www.jstor.org/stable/2684597

[43] Joanne Hinds and Adam N Joinson. 2018. What demographic attributes do our digital footprints reveal? A systematic review. *PloS one* 13, 11 (2018), e0207112.

[44] Thomas Hughes-Roberts. 2013. Privacy and Social Networks: Is Concern a Valid Indicator of Intention and Behaviour?. In *2013 International Conference on Social Computing*. IEEE, 909–912. https://doi.org/10.1109/SocialCom.2013.140

[45] Jane Im, Jill Dimond, Melody Berton, Una Lee, Katherine Mustelier, Mark S. Ackerman, and Eric Gilbert. 2021. Yes: Affirmative Consent as a Theoretical Framework for Understanding and Imagining Social Platforms. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 403, 18 pages. https://doi.org/10.1145/3411764.3445778

[46] Basileal Imana, Aleksandra Korolova, and John Heidemann. 2021. Auditing for Discrimination in Algorithms Delivering Job Ads. In *Proceedings of the Web Conference 2021* (Ljubljana, Slovenia) *(WWW '21)*. Association for Computing Machinery, New York, NY, USA, 3767–3778. https://doi.org/10.1145/3442381.3450077

[47] Carlos Jensen, Colin Potts, and Christian Jensen. 2005. Privacy practices of Internet users: Self-reports versus observed behavior. *International Journal of Human-Computer Studies* 63, 1-2 (2005), 203–227.

[48] Meike Kamp, Barbara Körffer, and Martin Meints. 2008. Profiling of Customers and Consumers-Customer Loyalty Programmes and Scoring Practices. In *Profiling the European Citizen: Cross-Disciplinary Perspectives*. Springer, 201–215.

[49] Munene Kanampiu and Mohd Anwar. 2019. Privacy Preferences vs . Privacy Settings: An Exploratory Facebook Study. In *Advances in Human Factors in Cybersecurity*, Tareq Z. Ahram and Denise Nicholson (Eds.). Springer International Publishing, Cham, 116–126.

[50] Ruogu Kang, Laura Dabbish, Nathaniel Fruchter, and Sara Kiesler. 2015. "My Data Just Goes Everywhere:" User Mental Models of the Internet and Implications for Privacy and Security. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*. USENIX Association, Ottawa, 39–52. https://www.usenix.org/conference/soups2015/proceedings/presentation/kang

[51] Bart P Knijnenburg, Martijn C Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. 2012. Explaining the user experience of recommender systems. *User modeling and user-adapted interaction* 22 (2012), 441–504.

[52] Ahmet Baki Kocaballi, Shlomo Berkovsky, Juan C Quiroz, Liliana Laranjo, Huong Ly Tong, Dana Rezazadegan, Agustina Briatore, and Enrico Coiera. 2019. The personalization of conversational agents in health care: systematic review. *Journal of medical Internet research* 21, 11 (2019), e15360.

[53] Richie Koch. 2020. Horizon 2020 Framework Programme of the European Union. 2021. Cookies, the GDPR,andtheePrivacyDirective. https://gdpr.eu/cookies/

[54] Anastasia Kozyreva, Philipp Lorenz-Spreen, Ralph Hertwig, Stephan Lewandowsky, and Stefan M Herzog. 2021. Public attitudes towards algorithmic personalization and use of personal data online: Evidence from Germany, Great Britain, and the United States. *Humanities and Social Sciences Communications* 8, 1 (2021), 1–11.

[55] Martin Krzywinski, Naomi Altman, and Paul Blainey. 2014. Nested designs. *Nature Methods* 11, 10 (2014), 977–979.

[56] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. 2018. Alexa, Are You Listening? Privacy Perceptions, Concerns and Privacy-Seeking Behaviors with Smart Speakers. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 102 (nov 2018), 31 pages. https://doi.org/10.1145/3274371

[57] Haein Lee, Hyejin Park, and Jinwoo Kim. 2013. Why do people share their context information on Social Network Services? A qualitative study and an experimental study on users' behavior of balancing perceived benefit and risk. *International Journal of Human-Computer Studies* 71, 9 (2013), 862–877.

[58] Una Lee and Dann Toliver. 2017. Building Consentful Tech. http://www.consentfultech.io/wp-content/uploads/2019/10/Building-Consentful-Tech.pdf

[59] Pedro Giovanni Leon, Blase Ur, Yang Wang, Manya Sleeper, Rebecca Balebako, Richard Shay, Lujo Bauer, Mihai Christodorescu, and Lorrie Faith Cranor. 2013. What Matters to Users? Factors That Affect Users' Willingness to Share Information with Online Advertisers. In *Proceedings of the Ninth Symposium on Usable Privacy and Security* (Newcastle, United Kingdom) *(SOUPS '13)*. Association for Computing Machinery, New York, NY, USA, Article 7, 12 pages. https://doi.org/10.1145/2501604.2501611

[60] Danny Yen-Ting Liu, Kathryn Bartimote-Aufflick, Abelardo Pardo, and Adam J. Bridgeman. 2017. *Data-Driven Personalization of Student Learning Support in Higher Education*. Springer International Publishing, Cham, 143–169. https://doi.org/10.1007/978-3-319-52977-6_5

[61] Yu-li Liu, Wenjia Yan, Bo Hu, Zhuoyang Li, and Yik Ling Lai. 2022. Effects of personalization and source expertise on users' health beliefs and usage intention toward health chatbots: Evidence from an online experiment. *Digital Health* 8 (2022), 20552076221129718.

[62] Mary Madden, Michele Gilman, Karen Levy, and Alice Marwick. 2017. Privacy, poverty, and big data: A matrix of vulnerabilities for poor Americans. *Wash. UL Rev.* 95 (2017), 53.

[63] Mary Madden and Lee Rainie. 2015. *Americans' attitudes about privacy, security and surveillance*. Pew Research Center.

[64] Michelle Madejski, Maritza Johnson, and Steven M. Bellovin. 2012. A study of privacy settings errors in an online social network. In *2012 IEEE International Conference on Pervasive Computing and Communications Workshops*. IEEE, 340–345. https://doi.org/10.1109/PerComW.2012.6197507

[65] Nathan Malkin, David Wagner, and Serge Egelman. 2022. Runtime Permissions for Privacy in Proactive Intelligent Assistants. In *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*. USENIX Association, Boston, MA, 633–651. https://www.usenix.org/conference/soups2022/presentation/malkin

[66] Alessandro Mantelero. 2016. Personal data for decisional purposes in the age of analytics: From an individual to a collective dimension of data protection. *Computer Law & Security Review* 32, 2 (2016), 238–255. https://doi.org/10.1016/j.clsr.2016.01.014

[67] Adam A. Margolin, Ilya Nemenman, Katia Basso, Chris Wiggins, Gustavo Stolovitzky, Riccardo Dalla Favera, and Andrea Califano. 2006. ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context. *BMC Bioinformatics* 7, 1 (20 Mar 2006), S7. https://doi.org/10.1186/1471-2105-7-S1-S7

[68] Arunesh Mathur, Mihir Kshirsagar, and Jonathan Mayer. 2021. What Makes a Dark Pattern... Dark? Design Attributes, Normative Considerations, and Measurement Methods. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 360, 18 pages. https://doi.org/10.1145/3411764.3445610

[69] Peter Mayer, Yixin Zou, Florian Schaub, and Adam J. Aviv. 2021. "Now I'm a bit angry:" Individuals' Awareness, Perception, and Responses to Data Breaches that Affected Them. In *30th USENIX Security Symposium (USENIX Security 21)*. USENIX Association, 393–410. https://www.usenix.org/conference/usenixsecurity21/presentation/mayer

[70] Aleecia McDonald and Lorrie Faith Cranor. 2010. Beliefs and behaviors: Internet users' understanding of behavioral advertising. *Tprc*.

[71] Aleecia M McDonald and Lorrie Faith Cranor. 2008. The cost of reading privacy policies. *Isjlp* 4 (2008), 543.

[72] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2021. A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR)* 54, 6 (2021), 1–35.

[73] George R Milne, George Pettinico, Fatima M Hajjat, and Ereni Markos. 2017. Information sensitivity typology: Mapping the degree and type of risk consumers perceive in personal data sharing. *Journal of Consumer Affairs* 51, 1 (2017), 133–161.

[74] Caroline Lancelot Miltgen and H. Jeff Smith. 2019. Falsifying and withholding: exploring individuals' contextual privacy-related decision-making. *Information & Management* 56, 5 (2019), 696–717. https://doi.org/10.1016/j.im.2018.11.004

[75] Josef Nguyen and Bonnie Ruberg. 2020. Challenges of Designing Consent: Consent Mechanics in Video Games as Models for Interactive User Agency. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376827

[76] Helen Nissenbaum. 2011. A contextual approach to privacy online. *Daedalus* 140, 4 (2011), 32–48.

[77] Patricia A Norberg, Daniel R Horne, and David A Horne. 2007. The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of consumer affairs* 41, 1 (2007), 100–126.

[78] Midas Nouwens, Ilaria Liccardi, Michael Veale, David Karger, and Lalana Kagal. 2020. Dark Patterns after the GDPR: Scraping Consent Pop-Ups and Demonstrating Their Influence. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376321

[79] Jonathan A Obar and Anne Oeldorf-Hirsch. 2020. The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services. *Information, Communication & Society* 23, 1 (2020), 128–147.

[80] Judith S. Olson, Jonathan Grudin, and Eric Horvitz. 2005. A Study of Preferences for Sharing and Privacy. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems* (Portland, OR, USA) *(CHI EA '05)*. Association for Computing Machinery, New York, NY, USA, 1985–1988. https://doi.org/10.1145/1056808.1057073

[81] Yong Jin Park. 2013. Digital literacy and privacy behavior online. *Communication research* 40, 2 (2013), 215–236.

[82] Frank Pasquale. 2015. *The black box society: The secret algorithms that control money and information*. Harvard University Press.

[83] Emilee Rader. 2014. Awareness of Behavioral Tracking and Information Privacy Concern in Facebook and Google. In *Proceedings of the Tenth USENIX Conference on Usable Privacy and Security* (Menlo Park, CA) *(SOUPS '14)*. USENIX Association, USA, 51–67.

[84] Emilee Rader and Rebecca Gray. 2015. Understanding User Beliefs About Algorithmic Curation in the Facebook News Feed. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 173–182. https://doi.org/10.1145/2702123.2702174

[85] Emilee Rader, Samantha Hautea, and Anjali Munasinghe. 2020. "I Have a Narrow Thought Process": Constraints on Explanations Connecting Inferences and Self-Perceptions. In *Proceedings of the Sixteenth USENIX Conference on Usable Privacy and Security (SOUPS'20)*. USENIX Association, USA, Article 24, 32 pages.

[86] Lee Rainie, Sara Kiesler, Ruogu Kang, Mary Madden, Maeve Duggan, Stephanie Brown, and Laura Dabbish. 2013. Anonymity, privacy, and security online.

[87] Kopo M. Ramokapane, Gaurav Misra, Jose Such, and Sören Preibusch. 2021. Truth or Dare: Understanding and Predicting How Users Lie and Provide Untruthful Data Online. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 557, 15 pages. https://doi.org/10.1145/3411764.3445625

[88] Joel R Reidenberg, Travis Breaux, Lorrie Faith Cranor, Brian French, Amanda Grannis, James T Graves, Fei Liu, Aleecia McDonald, Thomas B Norton, and Rohan Ramanath. 2015. Disagreeable privacy policies: Mismatches between meaning and users' understanding. *Berkeley Tech. LJ* 30 (2015), 39.

[89] Arjun Roy, Jan Horstmann, and Eirini Ntoutsi. 2023. Multi-dimensional Discrimination in Law and Machine Learning - A Comparative Overview. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) *(FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 89–100. https://doi.org/10.1145/3593013.3593979

[90] Shruti Sannon, Natalya N. Bazarova, and Dan Cosley. 2018. Privacy Lies: Understanding How, When, and Why People Lie to Protect Their Privacy in Multiple Online Contexts. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173626

[91] Shruti Sannon, Brett Stoll, Dominic DiFranzo, Malte F. Jung, and Natalya N. Bazarova. 2020. "I Just Shared Your Responses": Extending Communication Privacy Management Theory to Interactions with Conversational Agents. *Proc. ACM Hum.-Comput. Interact.* 4, GROUP, Article 08 (jan 2020), 18 pages. https://doi.org/10.1145/3375188

[92] Florian Schaub, Rebecca Balebako, and Lorrie Faith Cranor. 2017. Designing effective privacy notices and controls. *IEEE Internet Computing* 21, 3 (2017), 70–77.

[93] Florian Schaub, Rebecca Balebako, Adam L. Durity, and Lorrie Faith Cranor. 2015. A Design Space for Effective Privacy Notices. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*. USENIX Association, Ottawa, 1–17. https://www.usenix.org/conference/soups2015/proceedings/presentation/schaub

[94] Florian Schaub and Lorrie Faith Cranor. 2020. Usable and useful privacy interfaces. *An Introduction to Privacy for Technology Professionals* (2020), 176–299.

[95] Marco Scutari. 2010. Learning Bayesian Networks with the bnlearn R Package. *Journal of Statistical Software* 35, 3 (2010), 1–22. https://doi.org/10.18637/jss.v035.i03

[96] John S. Seberger, Marissel Llavore, Nicholas Nye Wyant, Irina Shklovski, and Sameer Patil. 2021. Empowering Resignation: There's an App for That. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 552, 18 pages. https://doi.org/10.1145/3411764.3445293

[97] Ruth G. Shaw and Thomas Mitchell-Olds. 1993. Anova for Unbalanced Data: An Overview. *Ecology* 74, 6 (1993), 1638–1645. https://doi.org/10.2307/1939922 arXiv:https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.2307/1939922

[98] Than Htut Soe, Oda Elise Nordberg, Frode Guribye, and Marija Slavkovik. 2020. Circumvention by Design - Dark Patterns in Cookie Consent for Online News Outlets. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society* (Tallinn, Estonia) *(NordiCHI '20)*. Association for Computing Machinery, New York, NY, USA, Article 19, 12 pages. https://doi.org/10.1145/3419249.3420132

[99] Daniel J Solove. 2012. Introduction: Privacy self-management and the consent dilemma. *Harv. L. Rev.* 126 (2012), 1880.

[100] Till Speicher, Muhammad Ali, Giridhari Venkatadri, Filipe Nunes Ribeiro, George Arvanitakis, Fabrício Benevenuto, Krishna P. Gummadi, Patrick Loiseau, and Alan Mislove. 2018. Potential for Discrimination in Online Targeted Advertising. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency (Proceedings of Machine Learning Research, Vol. 81)*, Sorelle A. Friedler and Christo Wilson (Eds.). PMLR, 5–19. https://proceedings.mlr.press/v81/speicher18a.html

[101] Sarah Spiekermann, Jens Grossklags, and Bettina Berendt. 2001. E-privacy in 2nd generation E-commerce: privacy preferences versus actual behavior. In *Proceedings of the 3rd ACM Conference on Electronic Commerce* (Tampa, Florida, USA) *(EC '01)*. Association for Computing Machinery, New York, NY, USA, 38–47. https://doi.org/10.1145/501158.501163

[102] Latanya Sweeney. 2013. Discrimination in Online Ad Delivery: Google Ads, Black Names and White Names, Racial Discrimination, and Click Advertising. *Queue* 11, 3 (mar 2013), 10–29. https://doi.org/10.1145/2460276.2460278

[103] Madiha Tabassum, Tomasz Kosinski, and Heather Richter Lipford. 2019. "I don't own the data": End User Perceptions of Smart Home Device Data Practices and Risks. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*. USENIX Association, Santa Clara, CA, 435–450. https://www.usenix.org/conference/soups2019/presentation/tabassum

[104] Xin Tan, Li Qin, Yongbeom Kim, and Jeffrey Hsu. 2012. Impact of privacy concern in social networking web sites. *Internet Research* 22, 2 (2012), 211–233.

[105] Songül Tolan, Marius Miron, Emilia Gómez, and Carlos Castillo. 2019. Why Machine Learning May Lead to Unfairness: Evidence from Risk Assessment for Juvenile Justice in Catalonia. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law* (Montreal, QC, Canada) *(ICAIL '19)*. Association for Computing Machinery, New York, NY, USA, 83–92. https://doi.org/10.1145/3322640.3326705

[106] Twillio. 2022. The State of Personalization 2022. https://segment.com/pdfs/State-of-Personalization-Report-Twilio-Segment-2022.pdf. Accessed: May 21, 2023.

[107] Alec Tyson and Emma Kikuchi. 2023. Growing public concern about the role of artificial intelligence in daily life.

[108] European Union. 2016. EuropeanParliament, 2016, Regulation (EU) 2016/679 of the European Parliament and of the Council. (2016). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679

[109] 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* (Washington, D.C.) *(SOUPS '12)*. Association for Computing Machinery, New York, NY, USA, Article 4, 15 pages. https://doi.org/10.1145/2335356.2335362

[110] Giridhari Venkatadri and Alan Mislove. 2020. On the Potential for Discrimination via Composition. In *Proceedings of the ACM Internet Measurement Conference* (Virtual Event, USA) *(IMC '20)*. Association for Computing Machinery, New York, NY, USA, 333–344. https://doi.org/10.1145/3419394.3423641

[111] M Vimalkumar, Sujeet Kumar Sharma, Jang Bahadur Singh, and Yogesh K Dwivedi. 2021. 'Okay google, what about my privacy?': User's privacy perceptions and acceptance of voice based digital assistants. *Computers in Human Behavior* 120 (2021), 106763.

[112] Ari Ezra Waldman. 2021. *Industry unbound: The inside story of privacy, data, and corporate power*. Cambridge University Press.

[113] Jeffrey Warshaw, Nina Taft, and Allison Woodruff. 2016. Intuitions, Analytics, and Killing Ants: Inference Literacy of High School-educated Adults in the US. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*. USENIX Association, Denver, CO, 271–285. https://www.usenix.org/conference/soups2016/technical-sessions/presentation/warshaw

[114] Miranda Wei, Madison Stamos, Sophie Veys, Nathan Reitinger, Justin Goodman, Margot Herman, Dorota Filipczuk, Ben Weinshel, Michelle L. Mazurek, and Blase Ur. 2020. What Twitter Knows: Characterizing Ad Targeting Practices, User Perceptions, and Ad Explanations Through Users' Own Twitter Data. In *29th USENIX Security Symposium (USENIX Security 20)*. USENIX Association, 145–162. https://www.usenix.org/conference/usenixsecurity20/presentation/wei

[115] Ben Weinshel, Miranda Wei, Mainack Mondal, Euirim Choi, Shawn Shan, Claire Dolin, Michelle L. Mazurek, and Blase Ur. 2019. Oh, the Places You've Been! User Reactions to Longitudinal Transparency About Third-Party Web Tracking and Inferencing. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security* (London, United Kingdom) *(CCS '19)*. Association for Computing Machinery, New York, NY, USA, 149–166. https://doi.org/10.1145/3319535.3363200

[116] Ben Weinshel, Miranda Wei, Mainack Mondal, Euirim Choi, Shawn Shan, Claire Dolin, Michelle L. Mazurek, and Blase Ur. 2019. Oh, the Places You've Been! User Reactions to Longitudinal Transparency About Third-Party Web Tracking and Inferencing. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security* (London, United Kingdom) *(CCS '19)*. Association for Computing Machinery, New York, NY, USA, 149–166. https://doi.org/10.1145/3319535.3363200

[117] Craig E Wills and Mihajlo Zeljkovic. 2011. A personalized approach to web privacy: awareness, attitudes and actions. *Information Management & Computer Security* 19, 1 (2011), 53–73.

[118] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 143–146. https://doi.org/10.1145/1978942.1978963

[119] Yuxi Wu, Sydney Bice, W. Keith Edwards, and Sauvik Das. 2023. The Slow Violence of Surveillance Capitalism: How Online Behavioral Advertising Harms People. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) *(FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 1826–1837. https://doi.org/10.1145/3593013.3594119

[120] Alice Xiang. 2020. Reconciling legal and technical approaches to algorithmic bias. *Tenn. L. Rev.* 88 (2020), 649.

[121] Yaxing Yao, Davide Lo Re, and Yang Wang. 2017. Folk Models of Online Behavioral Advertising. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 1957–1969. https://doi.org/10.1145/2998181.2998316

[122] Tomasz Zukowski and Irwin Brown. 2007. Examining the Influence of Demographic Factors on Internet Users' Information Privacy Concerns. In *Proceedings of the 2007 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries* (Port Elizabeth, South Africa) *(SAICSIT '07)*. Association for Computing Machinery, New York, NY, USA, 197–204. https://doi.org/10.1145/1292491.1292514