Normative and Non-Social Beliefs about Sensor Data: Implications for Collective Privacy Management

Emilee Rader
Michigan State University
emilee@msu.edu

Abstract

Sensors embedded in wearable and smart home devices collect data that can be used to infer sensitive, private details about people's lives. Privacy norms have been proposed as a foundation upon which people might coordinate to set and enforce preferences for acceptable or unacceptable data practices. Through a qualitative study, this research explored whether normative beliefs influenced participants' reactions to plausible but unexpected inferences that could be made from sensor data collected by everyday wearable and smart home devices. Some reactions were grounded in normative beliefs involving existing disclosure taboos, while others stigmatized the choice to limit one's use of technologies to preserve one's privacy. The visible nature of others' technology use contradicts individual concern about sensor data privacy, which may lead to an incorrect assumption that privacy is not important to other people. Findings suggest that this is a barrier to collective privacy management, and that awareness interventions focused on information about the beliefs of other users may be helpful for collective action related to data privacy.

1 Introduction

Sensors in wearable and smart home devices collect intimate information about people's bodies and activities in contexts that are usually considered to be very private. These data can be used to make new inferences about people that are difficult to anticipate and can be surprising, unsettling or harmful when used for unexpected purposes [18, 40, 41]. Privacy self-management, also called "notice and choice", is the established framework for data sharing rights and permissions [43]. Under this framework, an organization providing a sensor-enabled device and associated service sets its terms, and potential users must make a one-time, up-front, take-it-or-leave-it decision to consent to the terms or not. But when sensor data collection is automated, always-on and invisible, it is difficult to imagine how people can be making informed decisions about their preferences [29]. The consent decisions people make before ever using a technology may not reflect their beliefs and preferences once they have experienced using it [48]. In addition, as sensors embedded in everyday wearable and household devices allow service providers to amass more and more data, new inferences may become possible that were not at the time the user initially gave their consent [26]. For these reasons, privacy self-management fails as a mechanism for people to exert meaningful control over sensor data.

Because privacy self-management is so widespread, it is difficult to imagine what alternatives might look like. However, scholars have begun to suggest that a collective privacy management model based on norms for acceptable data collection, use and sharing might be a more natural and effective way for people to set boundaries for how information about them should be used [52]. Privacy norms are often described as a contextual factor that affects whether disclosure happens in a particular situation [37]. However, norms can also be thought of as a mechanism by which groups of people coordinate about behavior that is considered appropriate or inappropriate for the situation [7].

People adhere to norms for offline privacy-related behaviors [17, 39]. But, existing norms about private information might or might not influence people's beliefs and behaviors regarding the acceptability of sensor data collection and use. Anecdotally, it is possible to posit scenarios that support either position (that norms do or do not have an influence). For example, while people may believe that one should not physically sit outside someone else's home for hours at a time observing their comings and goings, many people install technologies such as doorbell cameras that record data about the behavior of neighbors and passers-by. In this example, a norm against spying on one's neighbors does not apply to adopting

USENIX Symposium on Usable Privacy and Security (SOUPS) 2022.
August 7–9, 2022, Boston, MA, United States.

a technology that effectively does the same thing.

The goal of this research was to investigate whether norms exist pertaining to the collection and use of sensor data. If normative beliefs play a role in determining acceptable and unacceptable sensor data practices, then it may be possible to design a method for collective data privacy management that relies on norm-based coordination among people. Sixty-five people who used activity trackers or voice assistants were interviewed about their own and others' reactions to hypothetical scenarios involving plausible but unexpected inferences made using sensor data collected by these technologies.

Participants' reactions to the scenarios demonstrated both normative beliefs and personal, non-social beliefs about the data and inferences presented in the scenarios. Normative beliefs involved existing disclosure taboos and also stigmatized the choice to limit one's use of technologies to preserve one's privacy. Personal beliefs focused on the desire to have control over data about oneself, the importance of awareness and consent, and the freedom of each individual to choose to use a technology or not according to their individual perceptions of how it could help them.

Norms arise where others' beliefs and behaviors are visible or known, and people can become aware of others' approval or disapproval. The choice to use a technology tends to be highly visible, whereas privacy-related concerns and motivations typically are not. The apparent contradiction between public behavior accepting data collection and private concern about it may lead people to an incorrect assumption that privacy is not important to others, and that engaging in privacy-preserving behavior is deviant. This may present a significant barrier to the development of collective privacy management strategies based on normative beliefs about sensor data. However, it suggests that awareness interventions focused on information about the beliefs of other users, rather than information about what sensor data are collected and shared, may be helpful for collective action related to data privacy.

2 Related Work

2.1 Social Norms

Social psychologists refer to two kinds of social norms: descriptive and injunctive. Descriptive norms are defined as "what is commonly done" [10]. They are beliefs about what others do, and arise through social comparison [21]. Injunctive norms are beliefs about "what is commonly approved/disapproved of" [10]. They are beliefs about what others believe, and are reinforced when specific feedback occurs in a given situation communicating to someone that their behavior violates the norm.

Social norms guide behaviors, but so do other types of beliefs, attitudes and values. This means that the presence of a norm cannot be determined by observing behavior alone—the same behavior might be caused by different kinds of beliefs. For example, Bicchieri [7] makes a distinction between independent but similar behaviors among a group of people that emerge from the needs and circumstances of a given situation (e.g., it is cold outside so everybody is wearing a heavy coat), and interdependent behaviors that arise through social influence. Interdependent behaviors can be caused by social imitation (e.g., everybody is wearing bow ties because they see everyone else doing it) which would be considered a descriptive norm. Or, interdependent behaviors can be caused by beliefs about the approval or disapproval of others (e.g., one should not ask someone else about how much money they make), which would be considered an injunctive norm because of the evaluative aspect.

To find out if there is a norm influencing behavior in a given situation, one must identify social beliefs and expectations that cause the behavior in question. If there's a correlation among the behaviors of a group of individuals, like everybody wearing a bow tie, the objective would be to find out whether this behavior serves some need or function and everyone is just coincidentally doing it, or if beliefs about others' beliefs or behaviors are causing the behavior to happen. Observations of actual behavior are important for identifying patterns, but not enough to tell what caused the behavior. One way to try to identify whether a behavior is norm-based is to identify factors that might have caused the correlation, and then ask questions about hypothetical situations that may or may not have occurred, to find out what people would do in those situations [7]. This makes it possible to discover whether social beliefs and expectations are associated with the behaviors, and thereby understand whether norms are at work.

2.2 Individual vs. Collective Privacy Management

Many conceptualizations of privacy treat it as an individual right, which means that individuals are responsible for controlling their own information according to their concerns and preferences [43]. Privacy is contextual, so it is difficult for people to know what their preferences for future contexts or inferences will be, based on the context in which they are making a privacy decision [37]. People can't make informed decisions when they're unaware of the consequences or don't have the expertise to figure them out [27]. And, there are too many different entities involved in collecting data about users for people to reason about them all individually. Solove [48] argues that the existing consent framework for *privacy self-management* is not working; while asking for consent makes data collection legally legitimate, it does not provide "meaningful control".

Collective privacy management is based on the idea that groups of people working together can coordinate to form and manage disclosure rules and boundaries [22, 46]. Much previous work on collective privacy management has focused on coordination among individuals about disclosure boundaries in social media. Multiple people may have different prefer-

ences about how a photo or other content should be seen and shared by others [6], and contextual factors like the nature of the relationship and network distance can play a role in negotiating and managing interpersonal boundaries [53]. Researchers have developed prototypes which, after individuals specify their privacy preferences, automatically merge preferences from multiple people to identify conflicts and propose or enforce boundary management solutions [4, 49, 50].

In addition to being used in research on interpersonal privacy, the phrase "collective privacy management" has also been used refer to policy and governance oriented approaches to managing data and information privacy. Sloan and Warner [46] argue that information privacy is a collective action problem, in which people have a common goal: to use technologies to meet their needs without disclosing information they don't want to disclose. However, the consent framework of privacy self-management does not support coordination between individuals or groups and the organizations collecting and using data about them.

Traditional grassroots organizing and activism may be one way for groups to argue for policies that would allow them to have more agency when choosing how their data may be collected and used [12]. Other research has explored ways to support people in coordinating with each other on privacy decisions. The coordination in these studies took the form of seeking advice from the community on privacy decisions [9], delegating consent for disclosure to trusted others [36], and presenting information to people faced with a privacy decision about others' privacy choices in similar situations [34].

In interpersonal privacy, disclosure rules and boundaries are often norm-based. People learn about appropriate and inappropriate disclosure behavior from others in their family or organizations they belong to, and form beliefs about what private information looks like and how it should be managed [39]. Those norms form part of the basis for collective interpersonal privacy management. It is difficult to envision what norms for data privacy look like, though, because agreeing upon conditions for the collection and use of digital data is typically treated as an invisible exchange between individuals and institutions. The goal of this study was to investigate whether norms exist pertaining to the collection and use of sensor data, and what specific normative beliefs might be present that could lay the groundwork for collective data privacy management.

2.3 Activity Trackers and Voice Assistants

Data privacy for sensor-based technologies is especially challenging because most sensors are by design invisible, embedded in everyday objects [19]. Data can be combined from multiple sensors, across points in time, and across multiple individuals or households to create inferences: new data points that cannot be directly collected from the environment by the sensors themselves, and are used to identify past patterns and predict future behavior [30].

When people initially purchase smart home technologies, often they are more focused on how they'll be able to use the features of the devices, and privacy concerns develop later [35]. Voice assistants, like Amazon's Alexa or the Google Home, include always-on microphones that can feel intrusive to some users [8]. Across multiple research studies, participants voiced concerns related to being unsure about what data was actually being collected about them, and worried about audio data being shared with third parties and either used for targeted advertising without their permission or used for unknown purposes [2, 20, 32]. Many participants in these studies talked about feeling powerless and unable to control the data that was collected about them [23, 32].

In contrast to voice assistant users, users of activity trackers in previous research tended to be unconcerned about privacy, because they believed fitness data (step counts, calories burned) are not sensitive. Step counts are perceived to be anonymous, and users are more concerned about looking good to others when sharing their fitness data than privacy threats they perceive to be unlikely [5]. In one study, the lack of privacy concern was attributed to the belief that it is not possible to accurately infer personal characteristics beyond fitness-related information from activity tracker data [51]. The only exception was if the activity tracker collected location data—this was seen as a potential privacy risk by some users [55].

This study builds on existing work about privacy concerns in these sensor-based devices, by focusing on the sensor data itself rather than perceptions and use of the technologies as a whole. In contrast to research focused on privacy risks and concerns related to expected uses of activity trackers (fitness tracking) and voice assistants (receiving and executing spoken requests in limited domain areas), this study involves hypothetical uses of the sensor data for inferences that could enable functionality beyond the intended purpose of the devices.

3 Method

3.1 Approach

Sixty-five semi-structured interviews were conducted in which participants were presented with hypothetical scenarios involving data collected by sensor-enabled technologies. The first round of 30 interviews focused on activity trackers, which are wearable devices equipped with accelerometers and other sensors that record data about the wearer's physical characteristics, like movements and heart rate. A second round of 35 interviews focused on voice assistants, such as smart speakers or integrated smartphone apps, which use microphones and speech recognition to accept questions and voice commands and respond by taking actions or providing information.

The hypothetical scenarios involved types of data that are typically collected by these technologies as part of their normal operation, so they would seem plausible to end users.

They described the data being used to make inferences that are not directly related to the typical usage scenarios for the two technologies. The scenarios were intended to prompt existing users of the technologies to imagine uses of the data they likely had not considered before, to elicit their initial reactions to these uses. Current or former users of these technologies were recruited to participate, so that participants could ground their reactions to the scenarios in their own experiences with the technologies, rather than relying on the interviewer's implicit or explicit framing of the technologies.

The interview questions asked about each scenario were based on a framework developed by Bicchieri [7] to help with identifying types of beliefs that guide people's choices and behaviors. For example, people may choose to act a certain way based on beliefs about what the outcome might be for them personally; beliefs about what they observe other people doing; or, they may hold beliefs about what one should do in a given situation. The difference between these is subtle but important. Consider the behavior of posting one's current salary on an online profile. Beliefs related to whether someone will do this or not might focus on possible retaliation from one's employer (non-social), seeing that others are/are not posting this information online (social), or anticipating that others will disapprove of posting one's salary online (normative) [11]. An intervention to encourage more people to be transparent about how much money they make would only be successful if it were tailored towards the type of beliefs preventing the behavior from occurring in the first place. The interview questions were designed using this framework in order to understand the types beliefs that underlie reactions to unfamiliar uses of sensor data, and to inform the analysis:

- *non-social beliefs* are based on one's own knowledge and experiences, and do not depend on others' beliefs and/or behavior
- *social beliefs* are based on one's expectations about how most others will behave in similar situations, and depend on observing others' behavior (descriptive norms)
- *normative beliefs* are based on one's beliefs about what others approve/disapprove of in similar situations (injunctive norms)

3.2 Interviews

Each interview began with background questions about the participant's use of the technology that was the focus of the interview, activity trackers or voice assistants. Most of each interview was spent presenting six hypothetical scenarios to the participant, one at a time, and asking questions to probe for reactions to each scenario. The scenario descriptions were brief, only a few sentences long. Each scenario mentioned both a type of data the device might collect (e.g., movements and location, content of recipes read aloud to the user by the device) and something the data might be used to infer (e.g., when the user went to the bathroom, how healthy the user's eating habits are). The scenarios did not present a rationale or motivation for the platform to do what the scenario described, nor for why the user would want to use the technology in the given scenario, so that participants were not biased or primed to understand the technology in the scenarios as serving a particular purpose. They also did not mention sharing the information in the scenario with third parties or other people.

The six scenarios were very different from each other, because participants' reactions were expected to vary according to their own beliefs and past usage of the technologies, and the interviews aimed to elicit a range of reactions from each participant. They were presented in the same order in each interview, and were designed to progress from more plausible inferences (closer to the intended purpose of the technology), to less plausible inferences. In pilot interviews, it was more difficult to gain participants' trust and build rapport when the scenarios with the least plausible inferences came first. Trust and rapport are necessary when asking about potential norm violations. This seemed an acceptable tradeoff for order effects for this investigation, which does not intend to make causal claims. See Appendix C for the text of the scenarios.

After introducing the first scenario, the interviewer began probing for participants' reactions by asking, "What are some different kinds of reactions people might have if [technology] could do this?" where [technology] was either activity trackers or voice assistants, referred to by the term the participant had used for the technology in the introductory part of the interview. The interviewer probed for specific examples and asked participants to explain terms and colloquialisms, and also used general prompts like "tell me more about that" to encourage participants to elaborate on their initial reactions. By asking about "different kinds of reactions people might have" the interviewer was encouraging the participant to consider not just their own reactions, but different ways they thought other people might react as well.

When the interviewer felt that the participant had nothing new to add about reactions to the scenario, they asked, "Do you feel like most people would think it is ok or not ok to use [technology] if it can know [information from scenario]?" and followed a similar strategy for probing for more detail. The third interview question asked about each scenario was, "How would you personally feel about using [technology] if it could know [information from scenario]?" This question was only asked if the participant had not already spoken about what they thought about the scenario. Once the participant had answered the three questions, the interviewer moved on to the next scenario.

The interview questions and follow-up prompts elicited reactions to each scenario in a neutral way, rather than framing the focus of the research as being about concern or privacy. The interviewer did not mention privacy or related ideas (e.g., surveillance, consent) unless the participant did first, which all participants did at some point during the interview. Likewise, the questions asked about the technology "knowing" the

information in the scenario instead of more precise terms like "infer", "calculate" or "detect" in order to avoid providing clues about how a system might do what was described in the scenario. See Appendix B for the interview questions.

3.3 Participants

Participants were recruited using a subject pool composed of volunteers from the community surrounding a large university in the midwest region of the United States, and by snowball sampling on social media to obtain greater geographic diversity in the sample. Close contacts of the researchers were ineligible, as were undergraduate students and people who reported having received formal training in computer science or IT (information technology).

In the first round of 30 interviews, participants were current or former users of wearable activity tracker devices (19 participants) or smartphone apps that tracked physical activity (11 participants). Eighty percent of participants in this round were women, and 60% came from snowball sampling. Participants ranged in age from 23 to 48 (M=33). Their self-reported occupations included stay at home mom, administrative assistant, graduate student, personal trainer, state government worker, sales associate, writer.

The second round of 35 interviews[1] focused on current or former users of voice assistants, described to potential participants as technologies similar to "Alexa, Hey Siri, or OK Google." Seventeen participants reported that they used Apple's Siri; the remaining used Google's voice assistant (13), Amazon Echo (6), Microsoft Cortana (3), and HTC Assistant (1). All of these except the Amazon Echo were apps on smartphones. About 30% of participants in the second round came from snowball sampling, and 46% of participants were women. Participants in the second round ranged in age from 20 to 72 (M=39), and their self-reported occupations included sports radio producer, chef, retired, small business owner, restaurant server, call center specialist, homemaker.

Recruiting for each round of data collection was conducted separately. At the conclusion of each interview, the interviewer created detailed memos describing emerging themes and similarities and differences across interviews. Recruiting continued until the majority of the reactions to the scenarios showed similar high-level themes to previous participants in that round. Overall descriptive statistics for both samples are presented in Appendix A. The interviews were conducted by telephone prior to the start of the COVID-19 pandemic, and ranged in length from 28 to 85 minutes (M=51). Each participant received a $25 Amazon.com gift card by email after the interview ended. This study was approved by the Michigan State University IRB.

[1]The voice assistant interview protocol initially used a scenario about inferring stress based on vocal pitch and speech patterns. However, the first several participants did not find this scenario plausible. A different scenario was used for the rest of the interviews, and five additional voice assistant interviews were conducted. The stress detector scenarios were not analyzed.

3.4 Analysis

Iterative qualitative analysis proceeded in several rounds [44]. First, the transcripts were coded for participant attributes like demographics, the type of activity tracker or voice assistant they used, etc. This round of coding also involved structural coding for which scenario was being discussed (Scenario 1-6) and which round of interview the transcript was from (activity tracker or voice assistant). This made it easier in later rounds of coding to identify which technology and scenario was the context for participant reactions.

Then, the transcripts were coded for inductive themes, focusing on statements indicating participants' beliefs and reasoning related to whether the data collection in the scenario was acceptable or unacceptable and why. Beliefs were loosely defined as thoughts and perceptions about what is true, based on personal knowledge and experiences [44]. For example:

- *acceptable because it doesn't seem harmful*: "And so if someone out there is tracking that about me, because I can't see what the harm is ultimately, maybe it's a little spooky, I don't know, but I feel like in this day and age, it's not even spooky anymore." (AT10, woman, 38, S4)[2]
- *unacceptable because being monitored is uncomfortable*: "That'll be kind of creepy. I don't know if I would like that. 'Cause it'll be almost like you were being watched, but through the microphone basically. I don't know if that's something that I would enjoy Siri knowing. I just don't think that's something that Siri needs to know about." (VA13, man, 39, S4)

Participants typically spoke about multiple beliefs related to the same scenario, even conflicting beliefs, as they considered the aspects of the scenario that came up while they thought about it and how others might react to it. In other words, participants could and often did make statements about both acceptable and unacceptable aspects of the scenarios, and not all of the beliefs they talked about were their own. The codes evolved through coding an initial set of about 10 transcripts across both rounds of interviews, and once the codes had stabilized the initial set was re-coded.

Then, another coding pass focused on just the segments of the transcripts coded with belief codes, and additional codes were applied that differentiated whether participants were talking about their own beliefs versus their beliefs about what other people believe. This coding pass also identified whether the beliefs evident in the transcript segments had either non-social, social or normative characteristics.

In the final stage of the analysis, the belief codes were grouped into several higher-level themes. Codes were combined that focused on similar reasons and explanations for

[2]Participants are referred to by ID number, gender, age and the scenario they were speaking about in the transcript excerpt. 'S4' stands for Scenario 4. 'AT' before the number indicates a participant in the activity tracker round of interviews; 'VA' indicates a participant in the voice assistant round. The full text of all scenarios can be found in Appendix C.

why the data collection and use in the scenario would be acceptable or not acceptable. These codes differentiated between statements focusing on privacy-relevant beliefs such as awareness, consent and control and those that did not.

3.5 Limitations

Participants' reactions to the hypothetical scenarios, and their beliefs about how others would react, should not be interpreted as accurate predictions about how they or others would behave if the scenarios were real. Privacy choices are context-dependent, and platforms and technologies often do not provide the options people would need to make choices according to their privacy beliefs and preferences. However, beliefs about privacy are important in their own right, because they are another factor that guides and constrains behavior. The goal of eliciting participants' reactions was to better understand normative influences on beliefs about appropriate versus inappropriate sensor data collection and use, to identify new opportunities for design and policy interventions that might help people better manage the privacy of their data.

The six scenarios used in this study were designed to seem plausible to participants and also to have potential privacy implications. The technologies involved, activity trackers and voice assistants, are both discretionary use technologies. This means that unlike smartphones or cars, these technologies are not necessary to support basic needs and activities. There may be beliefs and reactions related to non-discretionary technologies, or other uses of data from activity trackers and voice assistants not present in the scenarios, that were not elicited in this study due to the nature of the scenarios. In addition, if the scenarios had been presented in a different order, the specifics of participants' reactions may have varied. However, scenario order should not affect underlying beliefs.

Finally, this research used an opt-in convenience sample consisting of mostly white, highly educated people in the United States. The sample size, at 65 participants, is larger than many qualitative studies [31]. However, these findings should not be generalized to a more diverse population without being validated in a representative sample.

4 Findings

4.1 Norms about Private Information

Normative beliefs were present in many participants' reactions to multiple hypothetical scenarios. These beliefs focused on data collection and use about information and behaviors that participants said should be private or nobody else's business. Overall, 53 of the 65 participants (82% overall; 90% AT, 74% VA) across both rounds of the study had a reaction to at least one scenario that involved normative beliefs.

References to normative beliefs demonstrated an awareness of what others believe, like the following reactions from two participants to Scenario 3, about an activity tracker that could count how many times a person had used the bathroom. Here, AT17 (woman, 24, S3) described her expectation that nearly all other people would disapprove of the data collection and use in the scenario: "I think ninety-nine percent of people would say absolutely not. For no reason." Reactions involving normative beliefs also often had an evaluative component, like this belief described by AT26 (man, 28, S3): "Going to the bathroom's a personal thing, so it might just be a bit of a taboo subject."

In contrast, personal beliefs were typically spoken about in first person, as the participant's own belief rather than something everyone believes, e.g., "It just seems a little creepy to me, I don't know why, the phone knowing how often you oversleep" (VA16, woman, 56, S1). There were also instances where participants said they were unsure about what others would think, like the following from AT12 (woman, 39, S6): "I don't know. I would think that most people wouldn't care but I can also see why it would bother some people, but I guess I don't know about that." Statements like this were not considered to be examples of normative beliefs.

The most common reactions involving normative beliefs were about the hypothetical scenarios focused on bodily functions, like bathroom behavior and sleeping (24 participants, S2 and S3), about data collection in the home (24 participants, S4), and about inferring information about children (23 participants, S5). Forty-eight out of 65 participants (74% overall; 87% AT, 63% VA) described normative beliefs related to the use of the information in at least one of these three scenarios.

Many of the participants' reactions focused on how people in general feel that information about bathroom behavior is "personal" or "intimate" and is something one does not talk about with other people. Some spoke about how they felt like collecting this information would violate a taboo or be invasive of private space. For example, AT03 (woman, 32, S3) said, "People feel very personal about that [going to the bathroom], I don't think people would want anyone knowing that business." Participants had very little doubt or hesitation when they spoke about what others' reactions would be. They didn't equivocate—they were certain others would not like this. AT10 (woman, 31, S3) described it this way: "Oh, I think it would be outrageous. People would be outraged. Again, that's something that's very intimate, very personal." There was also an expectation that people would be angry if they found out this was being tracked without their knowledge. For example, "I would think people would just be, maybe, upset or angry that there would be information being kept on how many times you're going to the bathroom..." (AT21, woman, 40, S3).

Scenario 4 in both rounds of interviews involved the technology collecting data and making inferences about some aspect of the user's home environment. In the activity tracker interviews, the scenario involved the device making a map of the inside of the user's home while they wore it, and in the voice assistant interviews it involved doing voice detection

and counting the number of guests in the user's home. Participants' reactions to these scenarios centered on the idea that nobody would approve of this, because things that happen in one's home should be private. These participants talked about how if these inferences were being made, to most people it would feel like they were being "spied on" (VA02, man, 71, S4). Participant AT15 (woman, 36, S4) talked about how if her activity tracker did this, it would feel like being monitored—if the GPS and accelerometer data collected by the activity tracker were used for mapping rather than step counting, it would violate a norm about the home being private space: "I mean, if people wanted to know I could tell them, but personally people I don't think like to be monitored in their homes."

One scenario that was the same across both interviews involved data collected by the technology being used to infer whether the user had young children or not. Most of the reactions to this scenario invoked normative beliefs concerning protecting children from harm in general, and information about children more specifically. For example:

> "Oh man, I think that having young kids at home is such a huge personal line for people, that they... that would just probably be considered a huge, huge overreach, very intrusive, and posing a lot of security and personal safety issues." (AT16, woman, 29, S5)

Participants spoke with great confidence about this, even the participants who had no children themselves. For example, VA24, who did not have children, had this to say:

> "I think parents are bothered by everything involving people knowing things about their children they don't offer." (VA24, man, 27, S5)

A smaller number of participants (13 overall; 9 AT, 4 VA) talked about normative beliefs in response to other scenarios, particularly where it related to being healthy and hard-working as something people are supposed to do in order to be considered a good person. Most of these comments focused on the discomfort that comes with being evaluated negatively by others, and an expectation that the information in the scenario is something that people are often judged on. For example, in the following two examples a voice assistant participant and an activity tracker participant both spoke about beliefs about how people are supposed to behave in order to appear healthy:

> "Because especially for a woman, everybody thinks you're too fat or you're too thin. You're never perfect, and that's... If it's going to automatically evaluate you based on what you're cooking... Can we have one more thing not judging us?" (VA22, woman, 29, S3)

> "I think the majority of people would be afraid of being judged based on how many steps they do, or oversleeping an alarm... And we all accept that there's this basis of health that we're all supposed to maintain. There's this line that we all kind of say, ok, this is healthy living. Were you doing it or not? If we're not, we always feel guilty, and we always feel judged." (AT05, woman, 34, S1)

In the above excerpt, participant VA22 was reacting to Scenario 3 in the voice assistant round of interviews, which was about how a device with access to the user's recipes could read them aloud and assist them while they were cooking, but also make inferences about how healthy the user is based on characteristics of the meals they prepare. Her statement illustrates normative beliefs about womens' physical appearance as being related to her reaction to the scenario. Participant AT05 was reacting to Scenario 1 which was about an activity tracker that is worn to bed and counts how many times the user has overslept, and she felt that information could be used to categorize someone as lazy. These examples both illustrate very clearly the strong normative beliefs about how others approve or disapprove of people based on these characteristics.

Privacy theory considers norms to be part of the contextual factors that are important for people choosing whether or not to disclose private information [37, 39]. The findings in this section show that normative beliefs about the use of information about certain behaviors and contexts were part of participants' reactions to the scenarios. Because these norms (intimate behavior, home as private space, protecting children) are not specific to the digital context, it may seem obvious that normative beliefs about private information would apply to situations where technology is the observer of the information, not a person. However, it is also reasonable to hypothesize that people might feel like it is acceptable for their wearable devices or voice assistants to collect this information if it were not visible to other people or stayed on the device, or if the data were anonymous. The scenarios said nothing about whether the inferences would be shared or de-identified, so it is somewhat surprising that normative beliefs were present.

4.2 Norms about Privacy-Preserving Behaviors

In addition to norms about private information, there is also evidence in the data that norms exist regarding privacy-preserving behaviors, such as limiting one's use of technologies to restrict or prevent data collection about oneself. However, this evidence came in the form of normative beliefs that stigmatize concerns about data collection, and behaviors such as using less modern and sophisticated technologies (e.g., a flip phone; AT20, VA07, VA09, VA32) due to privacy concern.

A stigma is a strong sense of disapproval [38]. Stigmas often come about as punishment from a group for violating norms or deviating from accepted practices. To understand whether a stigma against privacy-preserving behaviors exists, first it is necessary to understand what people believe about normal, accepted practices related to the collection and use of digital data about themselves.

4.2.1 Data Collection is an Unavoidable Fact of Life

Thirty-three participants (51% overall; 43% AT, 57% VA) in this study believed, and also thought that other people believed, that digital data collection is an unavoidable fact of life. This is similar to the phenomenon of *digital resignation* described by Draper and Turow [14] and Seberger et al. [45]. These participants spoke about how it's not actually possible to choose not to have data collected about you—choosing to use technology is choosing to allow data collection. Participants said things like "This is the way the world is" (VA08); "I don't even know if I've agreed for them to pull out my information" (AT05); "by using the internet you're somewhat passively agreeing to be tracked" (VA02); "that's something that I feel is probably out of my control" (AT02). These participants did not seem happy about this, but rather unhappy and resigned.

VA05 (genderqueer, 24, S1) talked about it like a physical, physiological connection to their smartphone: "I mean, we're already so connected to our phones and now we have them monitoring our sleep and wake cycles like, plug me into my phone, we're the same being now." While this seemed uncomfortable for participants, there was also a sense of futility that made it difficult for them to rationalize objecting to it. As VA07 (man, 34, D5) said, "At some point it just becomes Google knows everything, and I have to deal with that if I... Either I use Google or I don't but they're gonna find out everything if I do."

Comments like these illustrate participants' beliefs about the data that has already been collected about them, and their reasoning about what data the technologies they use are capable of collecting about them. It was something they felt would be pointless to get upset over or do anything about, because it's already happened and is currently happening. For example, VA22 (woman, 39, S1) summed this up well: "I think people who use [voice assistants] are probably okay with it, 'cause they're already doing it. They're already doing things that are collected."

In addition, 22 participants (34% overall; 27% AT, 40% VA) talked about how they believed that using these systems indicates one must have consented, and that consent means people must be aware of the data collection and use practices (they "knew what they were getting into," AT15; or "knew about it going in," AT26). This was despite the fact that the participants themselves admitted not reading terms of use and privacy policies. These participants talked about data collection as an inevitable part of using technology, and used language that had connotations of defeat ("give up [information]," AT02), coercion ("forced to go along," VA09), and surrender/loss ("sacrifice," AT19) to describe it. They talked about unwanted data collection as common knowledge—something everyone knows is part of using technology and cannot be avoided ("you accept certain types of information being tracked," AT10). They rationalized uses of the data for purposes that were separate and unrelated to providing the service that was their reason for using the technology. But, at the same time, they said it was something that most people are not concerned about because, after all, they chose to use the technology ("Well you either buy the iPhone or you don't buy the iPhone," VA09).

These comments voice a belief that if a person chooses to use a technology or service, they are implicitly agreeing to everything that it does. For example, VA17 (man, 28, S4) said, "But in this theoretical scenario, I'm sure that probably the user has accepted via the application, the [voice assistant] or whatever, to do this sort of thing." This belief, consistent with the notice and choice framework, places the responsibility squarely on the user to know everything the technology is collecting and using. Participant VA09 described this well:

> "If you wanna use technology, I think that you have to accept the fact that you're gonna have data collected on you that you might not want to be collected on you." (VA09, man, 23, S3)

Two things are important about this for understanding whether a stigma against privacy-preserving behaviors exists. First, participants expressed personal beliefs that data collection is commonplace and inevitable, and they use these technologies anyway. And second, they believe that other people believe this as well. In other words, participants talked about social beliefs that most people accept and are fine with sensor data collection, and have consented to it. They know that choice or consent is required to use these technologies, and believe that the choosing to use them makes the individual responsible for what comes after.

4.2.2 Objecting to Data Collection Sounds "Crazy"

Twenty-three participants (35% overall; 17% AT, 51% VA) commented that people who are concerned enough about privacy that they believe technology is harmful, feel surveilled all the time, or are focused on other harms due to lack of privacy are crazy and/or paranoid. These comments arose when participants were asked about how they thought others would react to one of the scenarios, and frequently followed immediately after a statement about the participant's own desire to protect some aspect of the information about themselves in which they distanced themselves from that desire.

Paranoia is a delusional state in which a person is excessively suspicious about being targeted for harm by others without evidence that this is happening [42]. Previous research has also found that people perceive others who might use encryption tools as 'paranoid' [15, 54]. By associating this state with people who want to preserve their privacy, these participants indicated that they believe being concerned about privacy is at some level irrational and deviant. For example, AT25 described herself as "a little paranoid" because she "[doesn't] think that people need to know exactly what I'm doing every minute of every day." VA05 said that not

wanting Google to "know all these things about me" sounds "really paranoid." VA22 said, "I don't think that you can do anything electronically without it possibly coming back to you at some point, other people finding out about it." But, then she distanced herself from that belief by subsequently saying, "I don't want to sound like a paranoid person." VA33 talked about turning off location services on his mobile device, but then also said about himself, "I'm not super paranoid"—twice. This indicates that he believes that turning off location services could be viewed by others as paranoid, and he wanted to make sure that the interviewer knew he wasn't one of those paranoid people.

Participants also described that an unfounded, unreasonable anticipation of harmful outcomes is something that paranoid people do. For example, when asked about how she thought people would react to the scenario of an activity tracker collecting data about the inside of one's home, AT08 (woman, 42, S4) said, "I guess it depends on how paranoid you are and [the crime rate] where you live." VA35 (man, 51, S1) described worrying about "somebody finding out that they've hit the alarm so many times" as paranoid. And, VA24 (man, 27, S5) said, "Apparently, parents as a demographic seem like a paranoid group of people to me" after thinking about how parents would view a system that could automatically infer whether or not a person has children, and the possible harmful uses of that information. This indicates participants believed that paranoia is related to thinking about the likelihood and severity of privacy-related harms, and that paranoid people believe negative outcomes are more likely than is reasonable.

Believing in "conspiracy theories" was also often discussed as something that people who take steps to preserve their privacy do. Participants described that people with these beliefs feel like the government and companies are watching them and scrutinizing their activities, and that this feeling is extreme and unreasonable, even crazy (e.g., "crazy conspiracy nuts," VA07). So whereas the non-conspiracist beliefs held by the participants conveyed understanding that the data is being collected, it was considered to be a conspiracy theory to believe that the government and/or companies are paying attention and using that information for surveillance. AT28 (woman, 24, S1) described "the conspiracy theories people" as "the people who refuse to own smart phones because they believe the government is tracking their every move and that if you have a smart phone, you're signing away your right to all privacy ever." VA04 (woman, 32, S2) talked about how people should be "more suspicious" of data collection, "because there are people and programs that do want that kind of information, maybe the government." But then she immediately distanced herself from those beliefs by saying "I'm not a conspiracy theorist," like the people who "are rebelling against technology" because they are suspicious of it. And, VA26 (woman, 32, S1) gave an example of a coworker who doesn't want to share fitness tracker step counts with anyone, and referred to beliefs like that as, "a little more conspiracy theorist."

The findings in this section show that, in addition to norms about not disclosing some types of information and social beliefs about how everyone uses technologies that collect data about users, participants also held normative beliefs related to what they saw as deviant privacy-preserving behavior. They described that the common, accepted practice they engage in and see everyone around them also engaging in is to use technologies that everybody knows are collecting data about them. And, they talked in the abstract about how once a person chooses to use a particular technology, they've agreed to whatever data collection and use will take place. They also labeled people who object to this as "crazy" or "paranoid." This indicates that a norm was evident in their reactions to the scenarios, supporting acceptance of data collection as an unpleasant consequence of using technology, and labeling those who visibly object as deviant.

4.3 Non-Social Beliefs about Control over Data

In addition to the normative beliefs already described, participants also expressed private, non-social beliefs that indicated they do personally care about privacy, and that being able to keep some information private is important to them. These comments from participants emphasized the idea that they want to be aware of any changes to how the technologies and services they use are handling their data. They don't want the technologies to start doing something different with the data behind the scenes, like using it for some of the things described in the scenarios, without letting them know about it. Overall, 52 participants (80% overall; 73% AT, 91% VA) made statements like this in response to at least one scenario. For example,

> "But if that's a possibility we do need to be made aware of that, it can't just start happening." (VA04, woman, 32, S3)

> "I think it's important because I guess I want to know what information is being shared and being gathered. Even if it's just the totality of what is being tracked." (AT16, woman, 29, S4)

As part of speaking about a desire for control over how data about them is collected and used, participants talked about how they would not like it if the technology in the scenario were to start doing something they did not expect with the data. The participants' expectations were based on what they used the system for and what they believed it was doing. As VA18 said (woman, 30, S4), "it could be a little bit more off-putting that it could be just collecting more information about how many other people are around." In essence, these participants were saying if they're not aware of it and can't anticipate that it would need to be collecting that data, then it would not be acceptable to them. For example,

> "If they have that just hidden in there, like what Facebook does with a whole bunch of stuff, then

> no, I don't think it would be okay, and I think most people would be opposed it, if there was something that they were just sneaking in there." (VA23, man, 64, S4)

Twenty-seven participants (42% overall; 40% AT, 43% VA) emphasized that it was important to them personally to feel like they have a choice about opting in to any functionality that involves doing something they perceive to be new with the data that is collected by the technology in the scenario. There was an expectation articulated by these participants that if the technology wanted to do what the scenario described, then it would need to seek the user's permission first. VA29 (man, 45, S3) said, "Well, I imagine that it would be my choice to turn on [voice assistant] for this particular purpose, not that it would randomly come on." And, AT15 (woman, 36, S6) said, "So, that consent would at least, if somebody wanted to use the information for something, that they need to be very clear what they want to use it for, how it's gonna be used."

Thirty-nine participants (60% overall; 50% AT, 69% VA) made comments focused on the idea that participants want to have some control over aspects of the data collection and use described in the scenarios. A lot of these comments had to do with being able to keep the data on the device and not send it elsewhere or make it visible to others. They wanted to be able to create boundaries such that the data would only be used for the purpose the participant wants to use the technology for. To describe the types of data collection and inferences participants wanted to prevent, they used words indicating a boundary being crossed or violated, like "overreach" (AT16), "intrusion" (VA03, VA10), and "invasive" (AT21, VA10, VA17). VA07 talked about wanting to make sure that the voice assistant was not able to "hold onto that information any longer than it needs to", and AT18 (woman, 25, S6) talked about how the scenario would be more acceptable to her if she was able to turn off parts of that functionality: "I would be very confident if that's something I could be in charge of."

Participants' private, non-social beliefs about privacy are an interesting contrast with the normative beliefs, described in the previous section, about the acceptance of data collection and inferences as being just an inevitable (if unpleasant) part of using digital technology. Participants' private beliefs show that they do value the ability to have control over data about themselves, and are unhappy that they cannot.

4.4 Non-Social Beliefs about Usefulness

Participants' first thoughts immediately after hearing each scenario were nearly always focused on how they personally might use the functionality described in the scenario, or how it could help other people—as long as the scenario did not violate an existing norm. Overall, 64 participants (98% overall; 97% AT, 100% VA) made a statement about how important the usefulness/helpfulness of the scenario was for determining whether the data collection and inferences described were acceptable or not. This echoes the findings of research such as Dinev and Hart [13] and the recent literature review by Gerber et al. [16] about tradeoffs between the potential benefits of disclosing information and foreseeable harms.

4.4.1 Usefulness as Necessary Condition

Thirty-three participants (51% overall; 50% AT, 49% VA) believed that having access to the information in one of the scenarios over time would help them identify a pattern in their lives and behavior or in the world around them. Knowing about the pattern would then allow them to make better decisions, to change their behavior, or it would allow the system to make predictions or suggestions that would help them with their specific situation (e.g., changing the alarm time if you overslept a lot, making food substitution suggestions if you were eating too much salt/sugar, etc.). Participant VA09 (man, 23, S3) described his idea about how the scenario could help him: "you could have [voice assistant] suggest certain changes to your diet that she's been tracking for however long and you can be like, wow, I haven't eaten a fruit in two weeks, I should add an apple in or something." Similarly, AT15 felt that a greater awareness of oversleeping would be beneficial:

> "I guess that would at least give me a heads up like, 'Okay, maybe I need to do something different,' or, 'What can I do different so that I don't oversleep in the future?' So, I think it would be a positive thing." (AT15, woman, 36, S1)

In contrast, if the participant couldn't imagine a way that the information would be useful, then they felt the scenario would not be ok with other people, and the participant would not like it either. Twenty-four participants (37% overall; 17% AT, 54% VA) talked about how a particular scenario would not be useful because they thought it was not possible for the technology to make accurate inferences of the kind described in the scenario. A majority of these comments came from voice assistant participants who did not believe that microphones in one's home or smartphone could be used to accurately detect potential crimes being committed. For example, VA04 (woman, 32, S6) said, "maybe I'm having an argument with my boyfriend and it thinks, oh, there's domestic violence here. But really we're just having an argument. I think the data might be a little corrupted or just not accurate."

4.4.2 Useful or Not? It Depends...

Forty-four participants (68% overall; 67% AT, 69% VA) said that whether other people would find the functionality in the scenario useful would vary based on their beliefs, desires, characteristics or circumstances. These participants had difficulty even speculating about others' reactions to particular scenarios (that didn't violate norms or taboos) without knowing more about the other person's personality, preferences or life circumstances. This was most common in relation to

Scenario 6, which in the activity tracker interviews was about inferring one's carbon footprint from movement data, and in the voice assistant interviews was about inferring how safe one's neighborhood is from ambient sounds. In the carbon footprint scenario, participants talked about how others' reactions would depend on their beliefs about climate change, and in the crime monitoring scenario it would depend on how safe or unsafe one's neighborhood is. For example:

> "I think it depends on the individual person. If there's somebody who wants to reduce their carbon footprint, if they're looking to kind of get an idea, like a snapshot, of what their activities that impact on the environment around them." (AT21, woman, 40, S6)

> "So I guess if you're in a safe neighborhood, you'd probably say, great. It's giving my neighborhood a positive ranking. But if you're in one of the bad neighborhoods, you probably wouldn't like it." (VA23, man, 64, S6)

Two-thirds of participants (40, 62% overall; 73% AT, 51% VA) said that whether a given scenario would be useful or not would depend on additional information about the situation or context of use that were not provided as part of the scenario. The scenarios described sensor data being used by the technology to make inferences, but didn't provide much background or motivation for those inferences to be made or how a person would use the inferences in their lives. As such, the scenarios didn't contain the information participants felt like they needed to understand how the inferences could help someone, and this meant they could not conclusively say what their own or others' reactions would be. So when asked, participants talked about the kinds of things they believed would affect people's assessment of the scenario. These comments often focused on characteristics of possible harms in the scenarios that people would prefer to avoid, or whether or not the information would be shared. For example, 15 participants spoke about how people would react badly if the information in the scenario were shared with others when the user didn't want it to be, and another 15 participants talked about harms in the form of data breaches, higher insurance rates, or loss of physical safety due to the information being known.

Finally, over one third of participants (37% overall; 27% AT, 46% VA) said they didn't know or couldn't say what other people would think about at least one scenario. This was explained as just not having any idea (16 participants, e.g., "I don't know. I don't know what other people think," VA03), or not being able to say whether more people would be ok with it or not ok with it (9 participants, e.g., "I think it'd be pretty mixed... So 50-50 really," VA05).

Speculation about whether a scenario would be useful or not was a universal reaction to the scenarios, and an important perspective for participants' own evaluations of whether the scenario would be acceptable to them or not. In addition, participants believed that other people would also find usefulness to be an important factor, so much so that for most participants, more details about the situation and context were required to make a reasonable guess about others' reactions (again, for scenarios that did not violate existing norms). This indicates that there is no collective set of social or normative beliefs about usefulness related to whether or not one should or should not use a technology that collects a certain kind of data. Rather, usefulness is left up to the individual to determine for themselves.

5 Discussion

Norms are a form of collective action, in that they represent the convergence of beliefs among a group of people regarding behaviors that are acceptable and unacceptable. Notice and choice (privacy self-management) is the opposite of collective action—it makes the individual solely responsible for understanding the data practices and consequences of using a technology before they've even tried it, and once they've consented, makes it their fault if something happens that they don't like [48]. A collective approach to data privacy management would provide a framework for coordination among technology users so that they can take action as a group to set rules and policies for the data collection and use practices of organizations and platforms [12].

This research investigated whether normative beliefs play a role in people's reactions to plausible but unexpected inferences based on sensor data from common wearable and smart home devices. If norms do influence whether a particular inference is judged to be acceptable or unacceptable, then it is possible that collective privacy management strategies could be designed based on that foundation.

Normative beliefs were evident in participants' reactions to the hypothetical scenarios presented to them in this study. Common norms about disclosure of intimate information and protecting children were part of participants' reasoning for deciding that some scenarios would be unacceptable to them, and to most people. They were also uncomfortable with the idea that their voice assistants and activity trackers could use data collected as part of the technologies' normal operation to generate new inferences without informing them about what the inferences were and how they would be used. The existence of normative beliefs about unacceptable uses of sensor data is encouraging for the prospect of collective data privacy management. However, the findings of this study also identified three significant barriers that stand in the way of governance approaches or group collective action in support of better sensor data privacy solutions.

5.1 Barriers to Collective Data Privacy Management

The first barrier arises due to non-social beliefs about usefulness, and individual choice. The only universal rubric for deciding whether a scenario would be acceptable or unacceptable was how useful the data and inferences in the scenarios

might be. Participants initially considered each scenario from this perspective, and believed it would be of the utmost importance to other people as well. They also believed that it is up to each person to decide whether to use a technology or not according to their own individual circumstances, and everyone who uses technology accepts data collection as part of this and knows what they are getting into. This echoes the logic of privacy self-management, and supports the interpretation of others' continued use of potentially invasive technologies as an endorsement of the data practices those technologies employ. It is a highly individualistic approach that does not provide much common ground across people for collective data privacy management.

The second barrier is a result of social beliefs about technology use, and the inevitability of data collection. Participants believed data collection is an unavoidable fact of life if one chooses to use technology. The choice to use a technology tends to produce visible results, but privacy concern tends to have less visible outcomes. Knowing that others are using these technologies, participants assumed that most people approve of the company or service provider's data collection and use practices, because if they didn't they wouldn't consent and would not be using it. This makes it seem like nobody else cares about privacy, and perpetuates the belief that others must be comfortable with the status quo. This is a barrier to collective privacy management because it creates a descriptive norm supporting the use of potentially invasive technologies, no matter what their data practices are.

The third barrier stems from normative beliefs disapproving of privacy-preserving behaviors. Taking steps to limit data collected about oneself was viewed by participants as deviant, and individuals who do so were labeled as crazy or paranoid. During the interviews, some participants were even concerned themselves about being labeled in this way due to their speculation about possible harms from loss of privacy. But, participants' own non-social beliefs about the importance of controlling data and inferences about them contradicted this norm. In other words, privately, they valued privacy, but publicly they saw everyone not valuing it and negatively judging those who take steps to protect it. This contradiction is strikingly similar to a phenomenon called *pluralistic ignorance* [33], which occurs when people engage in behaviors they privately do not believe in or approve of, but they do it anyway because they believe that everyone else approves of it and they don't want to appear deviant.

Under conditions of pluralistic ignorance, normative beliefs about others' behaviors related to data collection and use conflict with private discomfort about the status quo. And in fact, stigmatization of people who violate the norm is a common component of pluralistic ignorance situations, and is especially difficult to combat when trying to change a prevailing norm [7]. Since there is little visible evidence that others value privacy and disapprove of privacy-violating data collection, people feel isolated in their private beliefs and are unlikely to speak up or take action. Pluralistic ignorance makes it extremely risky for individuals to speak up and advocate for better data privacy options and solutions. This would make meaningful reform of the existing data privacy governance structure (notice and choice) quite difficult.

5.2 Towards Collective Data Privacy Management

Effective governance of data collection and use practices based on collective data privacy management seems unlikely, given the barriers described above. Non-social beliefs echo the logic of privacy self-management, and both social and normative beliefs exist that are essentially anti-privacy. However, participants still wanted control over their data and disapproved of some types of data collection and use. The normative beliefs supporting privacy identified in this study all apply to any type of information, not just sensor data and inferences. In other words, they were unrelated to externalities created by massive datasets and machine learning. People's concerns about the lack of control over the data collected about them are generally invisible to others, making it nearly impossible for new norms related to sensor data and inferences to form. For example, imagine a norm similar to the existing norm about protecting children, but instead disapproving of providing data to a platform that could be used to harm someone else. For such a norm to form, information about others' beliefs about this would need to be more widely available.

Most approaches to helping people gain more control over their data focus on ways to make platforms' data practices more transparent to end users. But, awareness interventions focused on information about the beliefs of other users and their privacy choices, rather than information about what sensor data are collected and shared, may be helpful for collective action related to data privacy. People who use sensor-based technologies need to know they are not alone in their privacy concerns. Even small changes to the current notice and choice framework may create an opportunity to weaken the perception that others do not value privacy. For example, in April 2021, Apple provided a new feature in iOS 14.5 called App Tracking Transparency. This feature allows iPhone users to opt out of app data tracking. According to tech news sources, 50-60% of iPhone owners have chosen to opt out as of February 2022 [1,24]. However, platforms do not routinely disclose this type of information to end users.

Ultimately, privacy itself seems to be at odds with collective action. Behaviors like not disclosing information or opting out of using certain technologies are less visible than disclosing or opting in. In addition, choice—individual refusal—is the only option people believe they, and others, have for exercising control. But often, choosing not to allow data collection isn't really an option at all. Without more visible evidence of others' privacy-preserving beliefs, choices and behaviors, collective privacy management is unlikely to succeed.

Acknowledgments

The activity tracker interviews and pilots for the voice assistant interviews were conducted by Janine Slaker. In addition, the BITLab research group at Michigan State University provided feedback on earlier stages of this research. This material is based upon work supported by the National Science Foundation under Grant No. CNS-1524296.

References

[1] How Apple's privacy push cost Meta $10bn. *The Economist*, February 3 2022. https://www.economist.com/the-economist-explains/2022/02/03/how-apples-privacy-push-cost-meta-10bn.

[2] Noura Abdi, Xiao Zhan, Kopo M Ramokapane, and Jose Such. Privacy Norms for Smart Home Personal Assistants. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pages 1–14, 2021.

[3] Phil Adams, Mashfiqui Rabbi, Tauhidur Rahman, Mark Matthews, Amy Voida, Geri Gay, Tanzeem Choudhury, and Stephen Voida. Towards personal stress informatics: comparing minimally invasive techniques for measuring daily stress in the wild. *PervasiveHealth*, pages 72 – 79, 2014.

[4] Khaled Alanezi and Shivakant Mishra. Incorporating individual and group privacy preferences in the internet of things. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–16, 2021.

[5] Abdulmajeed Alqhatani and Heather Richter Lipford. "There is nothing that I need to keep secret": Sharing Practices and Concerns of Wearable Fitness Data. In *Symposium on Usable Privacy and Security (SOUPS)*, 2019.

[6] Andrew Besmer and Heather Richter Lipford. Moving beyond untagging: Photo privacy in a tagged world. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1563–1572, 2010.

[7] Cristina Bicchieri. *Norms in the Wild*. Oxford University Press, 2016.

[8] George Chalhoub, Martin J Kraemer, Norbert Nthala, and Ivan Flechais. "It did not give me an option to decline": A Longitudinal Analysis of the User Experience of Security and Privacy in Smart Home Products. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pages 1–16, 2021.

[9] Chhaya Chouhan, Christy M LaPerriere, Zaina Aljallad, Jess Kropczynski, Heather Lipford, and Pamela J Wisniewski. Co-designing for Community Oversight: Helping People Make Privacy and Security Decisions Together. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–31, 2019.

[10] Robert B Cialdini, Linda J Demaine, Brad J Sagarin, Daniel W Barrett, Kelton Rhoads, and Patricia L Winter. Managing social norms for persuasive impact. *Social Influence*, 1(1):3–15, March 2006.

[11] Zoë B. Cullen and Ricardo Perez-Truglia. The Salary Taboo: Privacy Norms and the Diffusion of Information. Technical Report NBER Working Paper No. 25145, 2020.

[12] Sauvik Das, W. Keith Edwards, DeBrae Kennedy-Mayo, Peter Swire, and Yuxi Wu. Privacy for the People? Exploring Collective Action as a Mechanism to Shift Power to Consumers in End-User Privacy. *IEEE Security & Privacy*, 19(5):66–70, 2021.

[13] Tamara Dinev and Paul Hart. An extended privacy calculus model for e-commerce transactions. *Information Systems Research*, 17(1):61–80, 2006.

[14] Nora A Draper and Joseph Turow. The corporate cultivation of digital resignation. *New Media & Society*, 21(8):1824–1839, 2019.

[15] Shirley Gaw, Edward W. Felten, and Patricia Fernandez-Kelly. Secrecy, flagging, and paranoia: Adoption criteria in encrypted email. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 591–600, 2006.

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

[17] Erving Goffman. *Behavior in Public Places: Notes on the Social Organization of Gatherings*. The Free Press, New York, NY, 1966.

[18] Samantha Hautea, Anjali Munasinghe, and Emilee Rader. That's not me: Surprising algorithmic inferences. In *Poster presented at the 2020 Symposium on Usable Privacy and Security*, 2020.

[19] Jong-yi Hong, Eui-ho Suh, and Sung-Jin Kim. Context-aware systems: A literature review and classification. *Expert Systems With Applications*, 36(4):8509–8522, 2009.

[20] Yue Huang, Borke Obada-Obieh, and Konstantin Beznosov. Amazon vs. My Brother: How Users of Shared Smart Speakers Perceive and Cope with Privacy

Risks. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20*, pages 1–13, 2020.

[21] Kyle Irwin and Brent Simpson. Do Descriptive Norms Solve Social Dilemmas? Conformity and Contributions in Collective Action Groups. *Social Forces*, 91(3):1057–1084, February 2013.

[22] Haiyan Jia and Eric P.S. Baumer. Birds of a feather: Collective privacy of online social activist groups. *Computers & Security*, 115:102614, 2022.

[23] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. Alexa, are you listening? privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW), 2018.

[24] Kif Leswing. Apple's ad privacy change impact shows the power it wields over other industries. *CNBC*, November 13 2021. https://www.cnbc.com/2021/11/13/apples-privacy-changes-show-the-power-it-holds-over-other-industries.html.

[25] Ying Li, Jose D Contreras, and Luis J Salazar. Predicting Voice Elicited Emotions. The 21th ACM SIGKDD International Conference, pages 1969 – 1978, 2015.

[26] Heather Richter Lipford, Madiha Tabassum, Paritosh Bahirat, Yaxing Yao, and Bart P. Knijnenburg. Modern Socio-Technical Perspectives on Privacy. pages 233–264, 2021.

[27] Eden Litt and Eszter Hargittai. A bumpy ride on the information superhighway: Exploring turbulence online. *Computers in Human Behavior*, 36:520–529, 2014.

[28] Hong Lu, Mashfiqui Rabbi, Gokul Chittaranjan, Denise Frauendorfer, Marianne Schmid Mast, Andrew T Campbell, Daniel Gatica-Perez, and Tanzeem Choudhury. Stresssense: detecting stress in unconstrained acoustic environments using smartphones. *Proceedings of the 2012 ACM international joint conference on Pervasive and ubiquitous computing*, 2012.

[29] Ewa Luger and Tom Rodden. An informed view on consent for UbiComp. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pages 529–538, 2013.

[30] David Lyon. *Surveillance Studies: An Overview*. Polity Press, Cambridge, UK, 2007.

[31] Bryan Marshall, Peter Cardon, Amit Poddar, and Renee Fontenot. Does sample size matter in qualitative research?: A review of qualitative interviews in is research. *Journal of Computer Information Systems*, 54(1):11–22, 2013.

[32] Nicole Meng, Dilara Keküllüoğlu, and Kami Vaniea. Owning and Sharing: Privacy Perceptions of Smart Speaker Users. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):1–29, 2021.

[33] Dale T. Miller and Cathy McFarland. Pluralistic Ignorance: When Similarity is Interpreted as Dissimilarity. *Journal of Personality and Social Psychology*, 53(2):298–305, 1987.

[34] Pardis Emami Naeini, Martin Degeling, Lujo Bauer, Richard Chow, Lorrie Faith Cranor, Mohammad Reza Haghighat, and Heather Patterson. The Influence of Friends and Experts on Privacy Decision Making in IoT Scenarios. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):1 – 26, 11 2018.

[35] Pardis Emami Naeini, Henry Dixon, Yuvraj Agarwal, and Lorrie Faith Cranor. Exploring How Privacy and Security Factor into IoT Device Purchase Behavior. Proceedings of the CHI Conference on Human Factors in Computing Systems, pages 1 – 12, 2019.

[36] Bettina Nissen, Victoria Neumann, Mateusz Mikusz, Rory Gianni, Sarah Clinch, Chris Speed, and Nigel Davies. Should I Agree? Delegating Consent Decisions Beyond the Individual. Proceedings of the CHI Conference on Human Factors in Computing Systems, pages 1 – 13, 2019.

[37] Helen Nissenbaum. Privacy as Contextual Integrity. *Washington Law Review*, 79:119–158, 2004.

[38] Bernice A. Pescosolido and Jack K. Martin. The stigma complex. *Annual Review of Sociology*, 41(1):87–116, 2015.

[39] Sandra Petronio. *Boundaries of Privacy: Dialectics of Disclosure*. State University of New York Press, Albany, NY, 2002.

[40] President's Council of Advisors on Science and Technology. Big data and privacy: a technological perspective. Technical report, May 2014.

[41] Emilee Rader, Samantha Hautea, and Anjali Munasinghe. I have a narrow thought process: Constraints on explanations connecting inferences and self-perceptions. In *Symposium on Usable Privacy and Security*, 2020.

[42] Nichola J. Raihani and Vaughan Bell. An evolutionary perspective on paranoia. *Nature Human Behaviour*, 3(2):114–121, 2019.

[43] Priscilla M Regan. Privacy as a Common Good in the Digital World. *Information, Communication & Society*, 5(3):382–405, 2002.

[44] Johnny Saldaña. *The Coding Manual for Qualitative Researchers*. Third edition edition, 2016.

[45] John S Seberger, Marissel Llavore, Nicholas Nye Wyant, Irina Shklovski, and Sameer Patil. Empowering Resignation: There's an App for That. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pages 1–18, 2021.

[46] Robert H. Sloan and Richard Warner. Why are Norms Ignored? Collective Action and the Privacy Commons. *Available at SSRN: https://ssrn.com/abstract=3125832*, February 18, 2018.

[47] H. Jeff Smith, Sandra J. Milberg, and Sandra J. Burke. Information Privacy: Measuring Individuals' Concerns about Organizational Practices. *MISQ*, 20(2):167 – 196, 1996.

[48] Daniel J Solove. Introduction: Privacy self-management and the consent dilemma. *126 Harvard Law Review*, pages 1880–1903, 2013.

[49] Jose M. Such and Natalia Criado. Resolving multi-party privacy conflicts in social media. *IEEE Transactions on Knowledge and Data Engineering*, 28(7):1851–1863, 2016.

[50] Kurt Thomas, Chris Grier, and David M. Nicol. unFriendly: Multi-party Privacy Risks in Social Networks. In Mikhail J. Atallah and Nicholas J. Hopper, editors, *Privacy Enhancing Technologies*, pages 236–252, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

[51] Lev Velykoivanenko, Kavous Salehzadeh Niksirat, Noé Zufferey, Mathias Humbert, Kévin Huguenin, and Mauro Cherubini. Are those steps worth your privacy? fitness-tracker users' perceptions of privacy and utility. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 5(4), 2022.

[52] Richard Warner. Notice and choice must go: The collective control alternative. *SMU Science and Technology Law Review*, 23(2):173–198, 2020.

[53] Pamela Wisniewski, A.K.M Namul Islam, Heather Richter Lipford, and David C Wilson. Framing and measuring multi-dimensional interpersonal privacy preferences of social networking site users. In *Communications of the Association for Information Systems*, volume 38, 2016.

[54] Justin Wu and Daniel Zappala. When is a Tree Really a Truck? Exploring Mental Models of Encryption. In *Symposium on Usable Privacy and Security*, pages 1–16, 2018.

[55] Michael Zimmer, Priya Kumar, Jessica Vitak, Yuting Liao, and Katie Chamberlain Kritikos. 'There's nothing really they can do with this information': unpacking how users manage privacy boundaries for personal fitness information. *Information, Communication & Society*, 23(7):1020–1037, 2020.

Appendix

A Participant Descriptives

Two rounds of interviews were conducted. The first focused on activity trackers (e.g., Fitbit), and the second focused on voice assistants (e.g., Amazon Alexa, Google Assistant, Apple Siri). Recruiting for the second round of interviews commenced after data collection of the first round was completed. Participants were recruiting using a subject pool composed of volunteers from the community surrounding a large university in the midwest region of the United States, and by snowball sampling on social media to obtain greater geographic diversity in the sample. This appendix presents overall descriptive statistics for both samples.

Note that at the end of each interview, participants were asked to fill out a brief demographic questionnaire which included questions from the Collection and Unauthorized Secondary Use subscales of the concern for information privacy (CFIP) instrument by Smith, Millberg and Burke [47]. These data were not analyzed for this paper; descriptive statistics are presented here as background. The privacy concern questions are listed below. The instructions were: "Here are some statements about personal information. From the standpoint of personal privacy, please indicate the extent to which you, as an individual, agree or disagree with each statement." The 8 items were presented in random order to each participant, and were answered on a 5-point Likert scale where 1 was Strongly Disagree and 5 was Strongly Agree.

Collection Subscale

- It usually bothers me when companies ask me for personal information.
- When companies ask me for personal information, I sometimes think twice before providing it.
- It bothers me to give personal information to so many companies.
- I'm concerned that companies are collecting too much personal information about me.

Unauthorized Secondary Use Subscale

- Companies should not use personal information for any purpose unless it has been authorized by the individuals who provided the information.
- When people give personal information to a company for some reason, the company should never use the information for any other reason.
- Companies should never sell the personal information in their computer databases to other companies.
- Companies should never share personal information with other companies unless it has been authorized by the individuals who provided the information.

Age	
Mean	36
Min	20
Max	72
SD	12
Gender	
Man	24
Woman	40
Other	1
Collection Subscale	
Mean	3.72
Min	1.25
Max	5
SD	0.81
Secondary Use Subscale	
Mean	4.43
Min	2.25
Max	5
SD	4.75

Table 1: Demographics of the 65 participants (30 from round 1 and 35 from round 2). Two participants (one from each sample) did not complete all CFIP [47] subscale items, and were excluded from those descriptives.

B Semi-Structured Interview Questions

Warm-Up Both rounds of interviews began with about 10-15 minutes in the interviewer introduced the study and asked some warm-up questions. These questions focused on learning more about the specific technology the participant used, the language and terminology they used to refer to it, the situations in which they used it, and the kinds of things they used it for. The interviewer did not mention privacy in the introduction to the study, and only asked follow up questions about privacy if the participant spoke about it first.

Transition to Scenarios After the warm-up questions, the interviewer transitioned to the hypothetical scenarios. At this point, the interviewer said something resembling:

> "Now, we're going shift our focus a bit from how you use your [technology], to thinking about some hypothetical scenarios about [technology]. The scenarios are different things that all might be possible in the near future, using the different kinds of information that [technology] can collect. We'd like you to imagine that the scenario is something that can really happen.
>
> The scenarios are designed to stretch your imagination and get you to think about ways of using a [technology] that you may not be used to, and how information generated by using a [technology] might be collected and used in the future.
>
> What I'm most interested in is your impressions

and ideas about different ways people might react to each scenario. So I have some follow-up questions for each scenario related to that. Do you have any questions?"

Questions about Scenarios The following questions were asked about each scenario presented to the participants:

- What are some different kinds of reactions people might have if [technology] could do this?
- A few participants had a hard time getting started talking about reactions to the scenario. Questions like these were used to prompt them to begin speaking about what they were thinking.
    - Can you imagine someone who would or wouldn't mind [technology] knowing this kind of information?
    - Tell me more about what makes it hard to imagine the scenario.
    - Who were you thinking of that might react like that?
    - Why do you think they would react that way?
- Do you feel like most people would think it is ok or not ok to use [technology] if it can know [information from scenario]?
- (If the participant hadn't answered this question yet...) How would you personally feel about using [technology] if it could know [information from scenario]?

C Text of Hypothetical Scenarios

Each round of interviews included six hypothetical scenarios involving possible future uses of data that could be collected by the technology (activity trackers or voice assistants). The scenarios are purely speculative, designed to seem plausible, but for the most part probably not something these technologies were actually doing at the time the interviews were conducted. Table 2 on the next page presents the text of each scenario used in the study.

Scenario 1 (S1) and Scenario 5 (S5) are very similar. The other scenarios were necessarily somewhat different, as the two technologies were quite different from each other and collected different kinds of data. However, even where the scenarios were different there are still some parallels:

- Scenario 2 (S2): both versions involve something that very closely resembles an existing use case for the two technologies.
- Scenario 3 (S3): both versions involve information that could be used to infer something about the user's health.
- Scenario 4 (S4): both versions involve always-on monitoring some type of information about the user's home environment
- Scenario 6 (S6): both versions involve information that some users might perceive as being in the public interest, that must be collected about the user and then aggregated across a group to produce a ranking

Scenario 6 in both rounds of interviews was a little bit different in that for both technologies it was about a societal issue (carbon footprint, crime) that has interdependent consequences beyond individual users. In other words, one person's carbon footprint or criminal activity in or near their home affects other people in the community (i.e., the environment or property values) in ways that oversleeping or the data and inferences in the other scenarios do not.

S1	*Activity Tracker*: Imagine a wearable sensor device that a user wears to bed. Some people use their activity trackers this way already. This hypothetical device can use information about the user's movements and alarm settings on their smartphone to know how many times the user overslept last week. *Voice Assistant*: Imagine a person that uses [voice assistant] as an alarm clock, to set an alarm to wake them up in the morning. Some people already do this, actually. Asking [voice assistant] to set an alarm means that it could use information about how many times the user hit snooze in the morning, or how long the alarm goes off before the user shuts it off, to know how often the user overslept last week. *Sensor*: Accelerometer (activity tracker), alarm time, interactions with device to snooze or stop the alarm
S2	*Activity Tracker*: Imagine an app that can use information from a user's wearable sensor device to make a graph or chart of when the user was sitting down at his or her desk at work each day last week. *Voice Assistant*: Imagine that [voice assistant] can be activated accidentally based on hearing the wake word when the user didn't actually intend to issue a command. This might happen if a user says the wake word when talking to someone else, or even when an actor in a TV commercial says it. This could allow [voice assistant] to know the content of some of the user's conversations when they don't mean to talk to the device. *Sensor*: Accelerometer, GPS (activity tracker); Microphone (voice assistant)
S3	*Activity Tracker*: Imagine that instead of time spent sitting down in a location, a wearable sensor device could use information about a user's movements and location to count how many times [he or she] went to the bathroom yesterday. *Voice Assistant*: Imagine that it's possible to use [voice assistant] while preparing meals, to read recipes and provide cooking instructions. This means that it would have access to information about ingredients, cooking methods, and meals the user prepares, and could determine how healthy a person is based on his or her eating habits. *Sensor*: Accelerometer, GPS (activity tracker); Microphone (voice assistant)
S4	*Activity Tracker*: What if an app were able to use information from a wearable sensor device to observe something about the user's environment based on their movements and altitude, like how many levels/floors there are in the user's home? What are some reactions you think other people might have to a device that could know that? *Voice Assistant*: What if [voice assistant] were able to use information from past voice commands to observe something about the user's home environment, like how many different guests or visitors the user has over? This could happen based on analyzing things like vocal pitch and speaking patterns, or the number of different voices in the background when a command is spoken. *Sensor*: Altimeter (activity tracker); Microphone (voice assistant)
S5	*Activity Tracker*: Imagine that it is possible for a system that uses wearable sensors to know whether a user has young children at home or not. This could be possible based on information about the user's movements, and GPS locations of places they visit, like playgrounds and parks. *Voice Assistant*: Imagine that it's possible for [voice assistant] to figure out whether the user is a parent who has a baby or toddler at home? This could be possible based on the content of the commands issued to the system, or the vocal pitch of the user, especially if a child asks [voice assistant] questions or directs it to play music. *Sensor*: Accelerometer, GPS (activity tracker); Microphone (voice assistant)
S6	*Activity Tracker*: Imagine that a system could estimate a user's weekly carbon footprint, and rank it against the carbon footprint of other users in their area. A wearable sensor device that can detect a user's movements and identify [his or her] GPS location can use this information to figure out when the user is in a moving vehicle, and estimate the carbon footprint based on that. *Voice Assistant*: Imagine that [voice assistant] could estimate how safe or unsafe the user's neighborhood is, and rank it against the safety level of the neighborhoods of other users in their region. This could be possible if at the same time as it is listening for the wake word, it is also listening for the sound of gunshots inside the home or nearby. Information about whether there has been gunfire at a location could be used to make a ranked list of each property and average that across a neighborhood. *Sensor*: Accelerometer, GPS (activity tracker); Microphone (voice assistant)

Table 2: Text of the hypothetical scenarios read to participants. The text in [brackets] was replaced by the terminology the participant used during the interview to refer to the technology they had experience with. The first two voice assistant interviews (DA01 and DA02) used a different Scenario 3, about detecting stress based on vocal pitch and speech patterns. However, both participants strongly felt this scenario was not believable, so the scenario was revised for the remaining interviews. (Note that detecting stress levels from audio data is actually feasible [3, 25, 28].)