Is This Thing On?
Crowdsourcing Privacy Indicators for Ubiquitous Sensing Platforms

Serge Egelman[1,2], **Raghudeep Kannavara**[3], **Richard Chow**[4]
[1]University of California, Berkeley, CA, egelman@cs.berkeley.edu
[2]International Computer Science Institute, Berkeley, CA, egelman@icsi.berkeley.edu
[3]Intel, Hillsboro, OR, raghudeep.kannavara@intel.com
[4]Intel, Santa Clara, CA, richard.chow@intel.com

ABSTRACT
We are approaching an environment where ubiquitous computing devices will constantly accept input via audio and video channels: kiosks that determine demographic information of passersby, gesture controlled home entertainment systems and audio controlled wearable devices are just a few examples. To enforce the principle of least privilege, recent proposals have suggested technical approaches to limit third-party applications to receiving only the data they need, rather than entire audio or video streams. For users to make informed privacy decisions, applications will still need to communicate what data they are accessing and indicators will be needed to communicate this information. We performed several crowdsourcing experiments to examine how potential users might conceptualize and understand privacy indicators on ubiquitous sensing platforms.

Author Keywords
Crowdsourcing; Ubiquitous Computing; Privacy

ACM Classification Keywords
H.5.2. Information Interfaces and Presentation (e.g. HCI): User Interfaces; D.4.6. Operating Systems: Security and Protection

INTRODUCTION
Within a week of its release, eight thousand beta testers signed up to pay $1,500 each for Google Glass [33], while countless others were added to a waiting list. We have also seen a recent demand for "smart watches," devices worn on the wrist and paired with a smartphone, allowing the user to take calls, view messages, and run third-party applications. One such device, the Pebble, raised over $10 million from 85,000 individuals using a crowd funding campaign [28], clearly illustrating public demand for these devices. Microsoft's Xbox One features voice and gesture control, as well as camera-based user recognition [26]. Intel has also announced that by mid-2014, many laptops will be equipped with their "RealSense" camera, which can be used to recognize gestures and voice commands [16]. All of these platforms allow third-party application developers to design applications that take advantage of new interaction modalities.

These devices all have the ability to perform ubiquitous sensing, which means that they require microphones and cameras to continuously monitor the user's environment for input. The sudden demand for these devices suggests that they will soon become a part of everyday life, which will create new security and privacy challenges [18]. Research on smartphone privacy and security has shown that granting third-party applications access to sensor data can create rich user experiences, but it can also be used for things other than application functionality (e.g., advertising and analytics), which users find concerning and surprising [25]. Thus, ubiquitous sensing platforms should employ privacy indicators so that users can better understand and control how data inferred from audio and video channels may be used by applications.

Research on smartphone privacy has shown us that care needs to be taken with the design of these privacy indicators. For instance, the Android operating system attempts to make this privacy information available to users, but as Felt *et al.* showed, most users do not notice or understand these notifications [11]. Thus, it is important to thoroughly evaluate privacy indicators prior to their widespread deployment. At the same time, privacy indicators for ubiquitous sensing systems pose an interesting challenge because researchers cannot fully predict all the ways that data from audio and video sensors might be used in the future.

We document the creation and evaluation of a set of privacy indicators for ubiquitous sensing platforms. Because these technologies are not yet widely deployed, we examined how potential users would conceptualize these privacy indicators and what pitfalls we might be able to identify prior to these privacy issues becoming too entrenched. These indicators are meant to be as universally understood as possible, so we adopted an approach based on crowdsourcing. We collected 238 sketches from participants based on 14 ubiquitous sensing concepts. We performed thematic analyses and used the uncovered themes as templates for new privacy indicator designs. We then performed comprehension evaluations with

CHI 2015, April 18 - 23 2015, Seoul, Republic of Korea.
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-3145-6/15/04...$15.00.
http://dx.doi.org/10.1145/2702123.2702251

over 1,500 participants. We discuss how some privacy concepts were well-understood and are therefore not necessarily endemic to specific technologies, whereas other concepts may require the technologies to be in widespread use before they can be adequately communicated to end-users. Besides identifying future focus areas for risk communication surrounding ubiquitous sensing technologies, we present a new method for privacy indicator design based on crowdsourcing.

BACKGROUND

While research pertaining to privacy in ubiquitous computing systems has been performed for over two decades, there are still many open questions about how to integrate usable privacy controls into specific systems, because it is impossible to know the full functionality of future systems. At the same time, researchers have shown that the "crowd" is a useful resource for many different research areas, including usable security and design [19, 17], and can be used to rapidly evaluate new systems and designs using sample sizes that were previously unheard of. In this section, we discuss some of the related work as it pertains to privacy for ubiquitous computing systems and crowdsourcing-based evaluations.

Ubiquitous Sensing

In 1991, Marc Weiser predicted the rise of ubiquitous computing and warned that "hundreds of computers in every room, all capable of sensing people near them and linked by high-speed networks, have the potential to make totalitarianism up to now seem like sheerest anarchy" [36]. He observed that many of these concerns could be mitigated by integrating privacy mechanisms into the earliest stages of system designs.

Over the past two decades, in recognition of the growing ubiquitous computing trend, many researchers have offered suggestions for how these systems should be designed (e.g., [2, 9, 20, 24, 1]), including methods for addressing privacy concerns (e.g., [29, 13, 15, 30]). For instance, Bellotti and Sellen introduced a set of eleven criteria for the evaluation of privacy in ubiquitous computing environments that focus on giving users control over information capture [3]. Langheinrich recommended a set of six principles for system design based on privacy guidelines from the Organization for Economic Cooperation and Development (OECD) [23]. However, most of this research has focused on general design principles, rather than specific implementations.

To help users understand how applications access sensitive data, most smartphone platforms employ some sort of permission system. Felt *et al.* examined the Android permission system, which uses install-time dialogs to list the abilities that an application will have, and found that most users do not notice or understand these permission requests [11]. Enck *et al.* developed the TaintDroid system to allow users to examine how applications may be using personal data in realtime [10]. Others have designed system to limit the amount of data received by applications (e.g., [4, 14, 12]). However, most of this research has been conducted after the platforms were already in wide deployment, making it difficult for developers to radically alter their systems to support the researchers' recommendations. Thus, our work is motivated by the fact that ubiquitous sensing platforms are beginning to receive public demand, so there is a unique opportunity to solve privacy and security problems before they become entrenched.

Roesner *et al.* looked at privacy and security issues related to the types of ubiquitous sensing platforms that our research is aimed at benefitting [31]. D'Antoni *et al.* suggested ways in which platforms can be designed to mitigate some of these problems [7]. Specifically, they advocated supporting the principle of least privilege: applications should only be given access to the data that they need to function, and nothing more [32]. Thus, if an application needs to recognize a voice command, the operating system should parse the raw audio and pass the command to the application without granting access to the entire raw audio stream, as this information is personally identifiable and could be used for other purposes. For systems to be trustworthy, they need to communicate how and when applications are accessing potentially sensitive data.

Crowdsourcing

Crowdsourcing, the process of leveraging large groups of people to perform short tasks, has attracted a lot of attention as a research tool [21]. Researchers have explored crowdsourcing as a tool for gathering subjective judgments about various online media. Most famously, von Ahn *et al.* created the ESP game to get Internet users to agree on image labels [35]. Kumar *et al.* used crowdsourcing to classify varying website layouts [22]. Bernstein *et al.* created a word processor that uses crowdsourcing for on-demand editing [5].

Dow *et al.* showed how crowdsourcing can benefit the design process. In their study, they used crowdsourcing to get feedback about different types of banner advertisements [8]. We build on this work by using crowdsourcing as an evaluation mechanism for the design of various privacy indicators.

Perceptual Computing

The privacy indicators that we designed for this project were directly inspired by Intel's RealSense SDK [16]. The SDK implements the recommendations of D'Antoni *et al.* [7]: application developers are provided with a set of API methods for extracting information from audio and video sensors so that those applications can request only the data needed to function, rather than requesting raw audio or video. This has the benefit of limiting privacy-sensitive information, while also making it easier for applications to make use of rich sensor data (i.e., the platform does most of the processing and feature extraction so that developers do not need to). In essence, the platform enforces a permissions model and the role of the indicators is to communicate what type of data the application is accessing. Note that in this paper we do *not* examine participants' privacy sensitivities surrounding the data. Instead, we concentrate on the issue of whether users understand what data is being requested by applications, a prerequisite for making informed privacy decisions.

INITIAL COMPREHENSION RATES

When we began this project in August of 2013, Intel was in the process of internally developing a set of privacy indicators. These initial indicators were designed *ad hoc* by an Intel

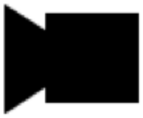
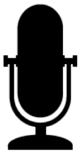
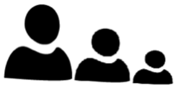
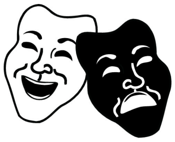
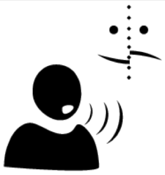
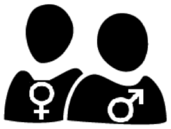
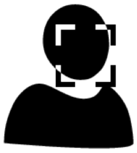
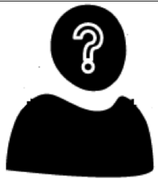
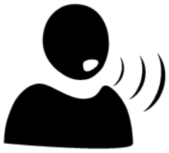
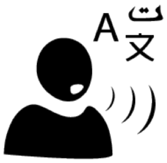
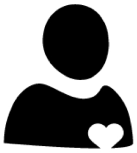
Video Recording	Audio Recording	Age Detection
		
96.5% of 57	94.4% of 89	8.7% of 46
Camera-Based Emotion Detection	Voice-Based Emotion Detection	Gender Detection
		
21.8% of 78	11.7% of 94	50.0% of 40
Face Detection	Face Recognition	Voice Command & Control
		
0.0% of 21	11.9% of 59	3.6% of 195
Language Detection	Heart Rate Detection	
		
16.9% of 77	14.7% of 190	

Table 1: Comprehension rates for the initial icons (the icon with the highest comprehension rate per concept is depicted).

UI designer. Our initial goal was to examine whether potential users would be able to understand the meaning of these icons, and if not, which ones needed to be improved to better convey their intended concepts. We received a total of 26 indicators representing 11 different concepts:

1. Video Recording (3 icons)
2. Audio Recording (2 icons)
3. Age Detection (4 icons)
4. Camera-Based Emotion Detection (3 icons)
5. Voice-Based Emotion Detection (2 icons)
6. Gender Detection (4 icons)
7. Face Detection (2 icon)
8. Face Recognition (3 icons)
9. Voice Command & Control (1 icon)
10. Language Detection (1 icon)
11. Heart Rate Detection (1 icon)

Open-Ended Survey

We recruited participants from Amazon's Mechanical Turk to participate in a survey on "camera icons." At the start of the task, we gave participants the following instructions:

> *Imagine in the future all computers (e.g., laptops, desktops, public displays/kiosks, etc.) will have cameras and microphones attached to them, which are always on so that applications can be controlled through gestures or spoken commands. Symbols or icons will be needed to indicate when a camera/microphone is recording, and whether an application has access to raw data (i.e., audio or video of the user) that can be used for any purpose, or if it can only access less-sensitive data that can only be used for limited purposes (e.g., determining the user's gender, a spoken command, the user's approximate age, etc.). In this survey, we will show you various icons that might be used for this purpose (i.e., these icons would appear in the designated area of the camera/microphone in the picture above). Your job is to describe what you think the camera/microphone is doing based on the icons that you are shown.*

Each page of the survey featured one of the icons followed by a text box asking, "based on this icon, what do you believe is being captured?" Each participant only viewed one random icon per concept, and we randomized the order of the concepts. We compensated participants $1.00 per survey submission and collected 208 responses in this manner. Our sample consisted of 56.7% females, an average age of 34 ($\sigma = 11.6$, range of 18-74), and 50.5% held a bachelor's degree or higher (23 held graduate degrees, including 4 doctorates).

Three researchers independently scored each response based on whether it matched the intended meaning for each icon. To measure inter-rater reliability, we calculated Fleiss' kappa ($\kappa = 0.757$). In total, we received descriptions for 2,181 icons, but for the remainder of our analysis, we only focus on the 1,853 (85%) for which all three raters were in agreement.

We were surprised at how similar the comprehension levels were for different icons depicting the same concepts. Of the 11 concepts that we examined, eight had multiple candidate icons. Thus, we performed Fisher's exact test to measure the comprehension rates between the highest and lowest ranked icons for each of these concepts. We performed eight tests in this manner and therefore applied the Bonferroni correction ($\alpha = 0.006$). Only in one case, recording video, was one icon significantly more intuitive than another for the same concept ($p < 0.004$). Regardless, for the remainder of our analysis we selected the icon with the highest comprehension rate in order to pair each concept with a single icon (Table 1).

Common Misunderstandings

By and large, video and audio icons were understood by almost everyone (97% and 94%, respectively). Therefore, we focused on the incorrect responses for the remaining nine icons to see if any patterns emerged, and whether these misconceptions could further guide the design process. Three independent coders performed a thematic analysis on the incorrect responses to the open-ended survey. They each identified the most common themes among the incorrect responses for each of these nine icons. We enumerated these themes and calculated the Intraclass Correlation Coefficient (ICC) to be approximately 0.94, indicating almost perfect agreement.

Age Detection

Of the 42 incorrect responses for the age detection icon, a majority of participants stated that the icon meant that a group of people were being captured on video: "multiple people in front of it," "it looks like a group of people are on the camera," and "the camera is capturing a group shot of people." These responses also mentioned families or video conferences. Averaging across all three coders, this theme accounted for 67% of the incorrect responses. Common among this theme was the misconception that the icon was meant to indicate multiple people, rather than multiple phases in an individual's life. Based on this, we believe that this icon could be improved by focusing on extreme differences in age, rather than simply using similar figures that only vary based on size.

Camera-Based Emotion Detection

Of the 61 incorrect responses to the camera-based emotion detection icon, a plurality of participants stated that they believed the icon indicated that a theatrical event was being captured (41%): "the camera is recording theatrical events or using a theater app," "this could signal making videos to put online for display," and "I think that a creative movie is being captured." We suspect that one reason for this confusion was that the drama masks used for the icon are a common theatrical symbol, and that this could potentially be corrected by focusing on the facial expressions and not using recognizable symbols with broader meanings.

Voice-Based Emotion Detection

Of the 83 incorrect responses to the voice-based emotion detection icon, a majority (63%) said that the icon indicated voice capture: "audio being captured" and "a person singing on camera." This suggests that participants focused on the figure speaking and overlooked (or were confused by) the emotion icon in the corner. We reasoned that applications employing emotion detection will receive the same data regardless of whether detection is performed with a camera or microphone, and therefore this distinction may be unnecessary, especially if it leads to confusion about what data an application is ultimately receiving.

Gender Detection

Of the 20 incorrect gender icon descriptions, most of them (85%) said that they believed that the icon was indicating *both* a male and female were present. An argument could be made that this is correct: if the device can infer that both genders are present, then the user understands that gender is being detected. This suggests that the gender aspect of the icon was clear, and that participants were confused by the silhouettes of two individuals next to each other. We felt the design could be improved by only showing one individual or using symbols.

Face Detection

None of our participants were able to correctly identify any of the face detection icons. The problem stems from distinguishing face *detection* with *recognition*; the former refers to determining whether an individual is present, whereas the letter refers to the identity of that individual. Of the 21 participants who viewed this icon, eighteen said it had to do with a picture being taken (86%), ten specifically mentioned that the box indicated the focus area: "this icon means the camera is capturing a visual, zooming in or focusing" and "the camera is determining where to focus." While it is technically correct to say that the camera is focusing on this area, this icon was supposed to indicate that only the user's presence would be shared with applications, and not her identity.

Face Recognition

Similar to the face detection icon, the face recognition icon had a very low rate of comprehension. Of the 52 incorrect responses, a plurality (37%) focused on concepts surrounding detection *errors*: "cannot tell who the person is...facial recognition did not work," "I believe my image is being recorded but the camera is having a hard time recognizing my face," and "the camera cannot discern who the person is." This suggests that much of the confusion was due to the question mark over the face; participants believed that this meant that the detection had failed, rather than that it was in progress.

Voice Command & Control

While the voice command icon also received a very low rate of correct responses in the open-ended survey, the incorrect responses were extremely similar: of the 188 incorrect responses, 83% said that the icon meant that either all audio or the user's voice was being captured. These responses were all incorrect, because the intent for the icon was to convey that only the name of a command would be accessible, rather than the user's voice, which is personally identifiable. This suggests that an improved icon might better demonstrate the command aspect, and downplay the speaking element.

Language Detection

We received a total of 64 incorrect responses for the language detection icon. A plurality of these (48%) had to do with the incorrect belief that voices and/or raw audio was being captured: "the microphone is capturing the user's voice," "the microphone is picking up sounds," and "the icon looks like your voice is being recorded." These errors were very similar to the errors seen with the voice command icon and therefore suggest that less emphasis should be placed on the speaker.

Heart Rate Detection

Finally, we received a total of 162 incorrect responses to the heart rate detection icon. Of these, a plurality (37%) had confused the icon with emotion detection. Specifically, they interpreted this icon to mean that the user is in love or is otherwise emotionally attached to someone or something: "this icon looks like this person really enjoys something," "this looks like romance," and "this seems to be indicating strong emotion, love in particular." One way of disambiguating the love aspect of the heart may be to indicate that it is beating, such as by superimposing an electrocardiogram (EKG).

Broadly speaking, the types of errors that we observed fell into four categories. First, many participants were confused by common symbols that they may have recognized in other contexts. For instance, in the heart rate detection icon, they

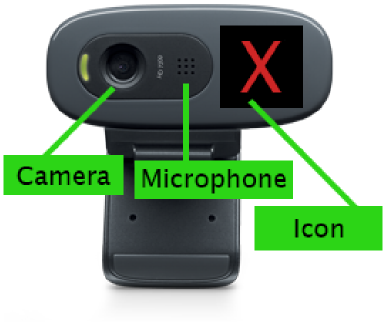


Figure 1: The picture used in our instructions so that participants could understand how the icons would be used.

assumed that the heart had to do with "love" or "emotions;" in the face recognition icon, they assumed that the question mark indicated that the computer was indicating an error; and in the emotion detection icon, they assumed that the drama masks had to do with drama or the theater. Second, many participants were confused when we attempted to describe both *what* and *how* data was being captured (e.g., camera-based emotion detection vs. voice-based emotion detection). Next, many participants found the depiction of multiple individuals in the gender and age icons to be misleading. Finally, several of the concepts were too similar to other concepts, for instance audio recording vs. voice command & control, or face detection vs. face recognition.

CROWDSOURCING THEMES

Our first experiment yielded insights into how our initial set of privacy indicators could be improved, however, it is possible that the comprehension rates were so low for some of the concepts because these indicators were the work of a single designer. We examined how potential users might choose to depict each of these concepts. We performed a second crowdsourcing experiment in which we described various ubiquitous sensing concepts and asked participants to draw their own icons to convey them. In this manner, we built a corpus of icons on which we performed thematic analysis to examine the common themes that each concept evoked.

Methodology

We posted a Mechanical Turk task with the following prompt:

> *Imagine in the future every computer has a camera and microphone that can be used to capture user input (e.g., spoken commands or hand gestures). This might include desktops, laptops, entertainment systems (TVs), and even public displays.*
>
> *To protect privacy, these devices will need to communicate when they are on and recording, and how this data will be used.*
>
> *In this task, we will describe a scenario and then you must design an icon to communicate how the recorded audio and/or video will be used. These icons will be displayed on the devices whenever an application is requesting data for a particular purpose.*

We then showed participants a picture of such a device (Figure 1) so that they could get a better idea of this icon's function. After reading these instructions, we presented participants with one of fourteen concepts,[1] drawn at random:

1. **Video Recording**: The video camera will record video of whoever is in front of it (i.e., applications will have access to all recorded video/images).
2. **Audio Recording**: The microphone will record audio of whoever is nearby (i.e., applications will have access to all recorded audio).
3. **Face-Based Age Detection**: The camera will only be used to determine the user's approximate age, and then will allow various applications to learn the user's age. However, no applications will have access to pictures or video of the user.
4. **Face-Based Emotion Detection**: The camera will take pictures to determine the user's approximate emotional state, and then allow various applications to learn the user's emotional state. However, no applications will have access to raw camera data (i.e., pictures or video of the user).
5. **Face-Based Gender Detection**: The camera will take pictures to determine the user's gender, and then allow various applications to learn the user's gender. However, no applications will have access to raw camera data (i.e., pictures or video of the user).
6. **Face Detection**: The camera will monitor whether a human being is in front of the computer and then notify applications when a user is present. However, applications will not have access to pictures or video of the user, nor will they learn the user's identity.
7. **Face Recognition**: The camera will take pictures to determine the identity of the individual in front of the camera, and then allow various applications to learn the identity of this user. However, no applications will have access to raw camera data (i.e., pictures or video of the user).
8. **Voice Command & Control**: The microphone will be used to recognize spoken commands, and then will share those commands with applications. However, no applications will have direct access to audio.
9. **Speech to Text**: The microphone will capture audio and convert it to text, and then allow various applications to access this text. However, no applications will have access to raw microphone data (i.e., audio from the user).
10. **Language Detection**: The microphone will be used to determine the language being spoken, and then will notify various applications of the user's language. However, no applications will have direct access to audio.
11. **Gesture Recognition**: The camera will take pictures to recognize gesture-based commands, and then allow various applications to access these commands. However, no applications will have access to raw camera data (i.e., pictures or video of the user).
12. **Voice-Based Emotion Detection**: The microphone will capture audio to determine the user's approximate emo-

[1] We included 3 new concepts supported by the latest RealSense platform that were not included with the previously-evaluated designer-created icons: speech to text, gesture detection, and eye tracking.

tional state, and then allow various applications to learn the user's approximate emotional state. However, no applications will have access to raw microphone data (i.e., audio from the user).

13. **Eye Tracking**: The camera will determine approximately where the user is looking on the screen, and then will share the coordinates with various applications. However, no applications will have access to raw camera data (i.e., pictures or video of the user).
14. **Heart Rate Monitor**: The camera and microphone will capture audio and video to determine the user's approximate heart rate, and then allow various applications to learn the user's heart rate. However, no applications will have access to raw camera or microphone data (i.e., video/images or audio of the user).

We included a use case for each concept to better convey why an application may want access to that data (and why it might be beneficial to the user). Alongside the concept description and use case, we provided participants with a sketching applet, as well as a box to explain how their drawing illustrates the given concept. We restricted our task to participants 18 years of age or older, but did not restrict it to a particular geographic area, in hopes of receiving submissions from a wide variety of individuals. We paid each participant $0.25.

Results

We received a total of 274 different sketches across all 14 concepts (an average of 17 sketches per concept). We discarded 36 (13%) sketches because the textual descriptions either contained gibberish or had nothing to do with the concepts, indicating that these participants did not take the task seriously. With the remaining 238 sketches, we performed a thematic analysis. Three of us independently made lists of themes that appeared across multiple sketches of the same concept. We then created a codebook based on these themes and two of us enumerated how many sketches embodied each of the coded themes. Based on our independent calculations of how frequently each theme occurred, we calculated the ICC to be 0.86, indicating almost perfect agreement. Our next step was to establish how many themes for each concept we would focus on in future experiments. After observing that for most concepts the top two themes were found in a majority of sketches, we focused on only the top two most prevalent themes in each concept. Table 2 depicts these themes and Figure 2 depicts some of participants' drawings.

Based on these themes, we made several observations. First, we were surprised that at least one of the themes for each concept appeared in our set of designer-provided icons:

- The video and audio icons embodied the camera and microphone, respectively.
- The age detection icon depicted child and adult.
- The emotion detection icon featured a happy and sad face.
- The gender detection icon featured male/female symbols.
- The face detection icon featured a framed face.
- The face recognition icon featured a face.
- The voice command icon featured a person speaking.
- The language detection icon featured a mouth with foreign characters.
- The voice-based emotion detection icon featured both sound waves and a happy and sad face.
- The heart rate monitor icon featured a heart.

Concept	**Top Themes**
Video Recording (19)	• camera (10) • recording light (7)
Audio Recording (25)	• microphone (16) • sound waves (14)
Face-Based Age Detection (16)	• child or adult (6) • child and adult (4)
Face-Based Emotion Detection (13)	• smiley face (9) • happy and sad face (6)
Face-Based Gender Detection (14)	• male/female symbols (7) • male/female figures (5)
Face Detection (16)	• face (15) • camera frame or crosshairs (5)
Face Recognition (16)	• face (14) • camera frame or crosshairs (7)
Voice Command & Control (13)	• person speaking (6) • sound waves (6)
Speech to Text (15)	• letters or text (11) • sound waves (7)
Language Detection (21)	• foreign characters/words (9) • mouth (5)
Gesture Recognition (11)	• hand (10) • waving motion (6)
Voice-Based Emotion Detection (14)	• happy or sad face (14) • sound waves (7)
Eye Tracking (25)	• eyes (24) • arrow or directional lines (8)
Heart Rate Monitor (20)	• heart (14) • EKG (11)

Table 2: The two most prevalent themes for each of the 14 concepts for which participants submitted sketches. The number after each concept reflects the total number of sketches, whereas the numbers after the themes reflect how many sketches reflected each theme (not mutually-exclusive).

We also observed that the exact same themes emerged for both the face detection and face recognition icons, which also occurred in our initial comprehension survey. This suggests that users may have a very difficult time differentiating these two concepts. This is especially concerning since they represent very different privacy concerns: an individual's identity vs. that *an* individual is present. Because we believe that the latter represents a much lower risk, we decided to eliminate the face detection concept from future evaluations.

Finally, we previously observed that between camera-based emotion detection and voice-based emotion detection, the same data would be accessed—the user's emotional state. Therefore, because it does not matter how that emotional state is determined, we merged these two concepts into a single indicator. This resulted in a set of 12 final concepts for which we made new privacy icons based on the themes in Table 2.

EVALUATION

Using the themes that we uncovered from participants' sketches, we created icons for each of the 12 concepts. For each concept, we created an average of six icons that em-

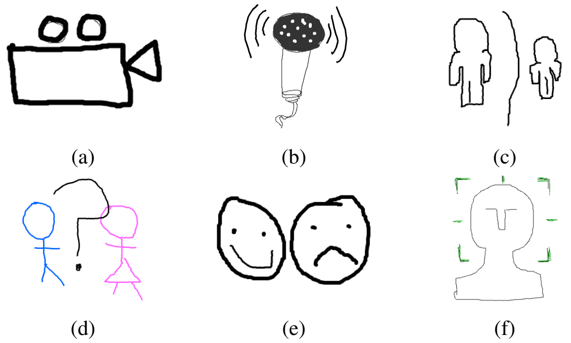


Figure 2: Crowdsourced drawings for six concepts: video recording (a), audio recording (b), age detection (c), gender detection (d), emotion detection (e), and face recognition (f).

bodied the themes in different ways. We based our icons on concrete concepts, following the findings of Curry *et al.* that concrete icons are initially much more understandable than abstract icons (e.g., line art) [6].

We iteratively deployed a survey 5 times, wherein participants defined an icon by choosing from a set of concepts. With each iteration, consisting of roughly 250 participants, we reduced the set of candidates until we were left with 1-2 icons per concept. Ultimately, we tested 71 different icons on 1,234 participants. We then performed a final evaluation to compare the designer-supplied icons (control condition) with our final set of crowdsourcing-inspired icons (experimental condition). In this section, we describe this final evaluation.

We presented participants with the following instructions:

> *Imagine in the future all computers (e.g., laptops, desktops, public displays/kiosks, etc.) will have cameras and microphones attached to them, which are always on so that applications can be controlled through gestures or spoken commands, as well as to learn things about their users. Symbols or icons will be needed to indicate what data is being collected about the user.*
>
> *In this survey, we will show you various icons that might be used for this purpose (i.e., these icons would appear in the designated area of the camera/microphone in the picture above). Your job is to describe what data you believe is being collected based on the icons that you are shown.*

On each of the 12 pages of the survey (one page per concept), we randomly displayed either a control condition icon or an experimental condition icon[2] and asked them, *based on this icon, what data do you believe is being collected?* We provided them with the following options, in random order:

- The language spoken
- A hand gesture
- Text of anything spoken (speech to text)
- Your approximate age
- Eye tracking data (where you are looking)
- Video of you
- A command (voice control)
- Your gender
- Your emotional state (mood)
- Audio of you
- Your face or identity (face recognition)
- Your heart rate

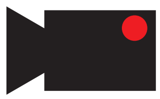
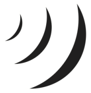
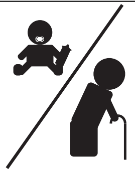
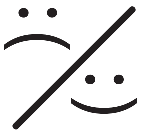
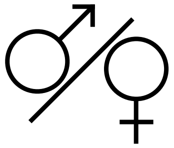
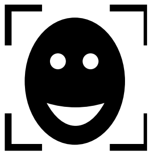
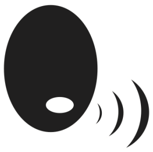
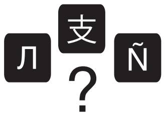
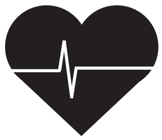
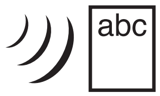
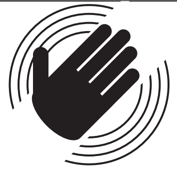
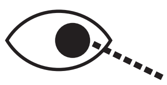
Video Recording	Audio Recording	Age Detection
 E: 84.7% of 138 C: 85.8% of 141 $p < 0.872$	 E: 68.8% of 144 C: 79.4% of 160 $p < 0.036$	 E: 86.3% of 160 C: 27.1% of 144 $p < 0.001$
Emotion Detection	Gender Detection	Face Recognition
 E: 91.2% of 148 C: 83.3% of 156 $p < 0.058$	 E: 96.2% of 104 C: 97.1% of 105 $p < 0.721$	 E: 74.0% of 104 C: 85.3% of 95 $p < 0.055$
Voice Command & Control	Language Detection	Heart Rate Detection
 E: 44.1% of 59 C: 42.6% of 54 $p < 1.000$	 E: 90.5% of 158 C: 67.6% of 145 $p < 0.001$	 E: 98.8% of 160 C: 90.9% of 143 $p < 0.002$
Speech to Text	Gesture Recognition	Eye Tracking
 E: 73.0% of 141	 E: 98.6% of 147	 E: 98.5% of 134

Table 3: Results of the final evaluation survey with the experimental icons displayed. The comprehension rates are listed below each icon for both the experimental (E) and control (C) conditions. Comprehension rates in the experimental condition were significantly higher for three concepts ($\alpha = 0.006$): age detection, language detection, and heart rate detection.

For each concept, we performed Fisher's exact test on the comprehension rates between the control icon and the experimental icon with the highest comprehension rate. To account for multiple testing (nine of the twelve concepts fea-

[2] In the final evaluation, we examined two different experimental icons for the gender detection and face recognition concepts, and three icons for the voice command icon.

tured designer-supplied control icons), we applied the Bonferroni correction ($\alpha = 0.006$). Our sample ($n = 304$) was 52.2% male (158 of 304), the average age was 33 years old ($\sigma = 11.3$, range of 18 to 68), and 43.1% held a bachelors degree or higher (23 participants held graduate degrees, including two doctorates). Thus, we believe our sample is representative of the U.S. online population.

Overall, we were surprised at how similar comprehension rates were between the two icon sets (Table 3): none of the control condition icons had significantly higher comprehension rates than the corresponding experimental condition icons. This indicates that the crowdsourced icons were no worse than the original set. More importantly, three experimental condition icons exhibited significantly higher rates of comprehensions than their control condition counterparts. Thus, for a third of the concepts—age detection ($p < 0.001$), language detection ($p < 0.001$), and heart rate detection ($p < 0.002$)—the icons that we created from crowdsourced themes were more intuitive than the initial icons.

DISCUSSION

In our initial experiment, we asked participants to provide open-ended explanations for what they believed various designer-drawn privacy indicators represented. In our last evaluation, we asked participants to perform a similar task, but instead selecting their responses from a multiple-choice list of possible explanations. As one might expect, the latter evaluation resulted in higher comprehension rates.[3] We believe that taken together, these two different ways of evaluating comprehension—open-ended vs. multiple-choice responses—represent lower and upper bounds, respectively.

As Moyes *et al.* observe, "if an icon is not guessable it is not necessarily an unsuccessful icon" [27]. They hypothesize that learnability through repeated exposure may bridge this gap. Because the technologies that we hope to influence are not yet widely available, it is unlikely that many subjects were familiar with their capabilities, which is likely to result in lower rates of comprehension—guessability—in response to the open-ended questions (than if subjects had familiarity with the devices). This may be why subjects exhibited much higher rates of comprehension for the video and audio recording indicators: technologies that perform these actions are already in widespread use, and therefore concrete representations of these concepts were recognizable.

Along these lines, we believe that one of the biggest as-yet-unsolved challenges that we faced with this work was distinguishing face recognition from face detection. The former is used to identify an individual user, whereas the latter is used to determine whether someone is present (without identifying them). In our initial experiments, the designer-created icons were unable to disambiguate these concepts. We observed that during the sketching task, the themes that we extracted from the drawings representing each concept were identical. Likewise, during the comprehension experiments, participants' responses to these two concepts were interchangeable. Due to this confusion, we ultimately decided to remove the face detection concept because we felt that the privacy concerns associated with it were minor in comparison to those associated with face recognition; as Thompson *et al.* recommend, indicators that represent very minor risk levels should be eliminated so as to prevent users from becoming habituated to indicators representing much more serious risks [34]. We suspect that subjects' lack of familiarity with these concepts may be responsible for their inability to distinguish them. Additionally, it is possible that these concepts are just so similar, that the only way to communicate the concepts will be through learned association. That is, this problem may only be solved by simply assigning an icon to each concept and then expecting subjects to learn each icon's meaning through repeated exposure.

Limitations

This study was not without its limitations. Specifically, we have several questions about the generalizability of our results. First, with the exception of the drawing task, all of the evaluations were performed on participants based in the U.S. Obviously, the systems that we hope to benefit through this research will be deployed globally. Therefore, more work needs to be done to examine whether these indicators are effective at communicating the concepts to an international audience. This may pose a challenge, as the evaluations need to be conducted in a language that is understood by all participants, which could potentially mean hundreds of variants of a single experiment in order to ensure that an indicator is universally recognized.

Another limitation of this study is that we only examined a single set of control icons. Our control icons were developed internally at Intel by a professional designer and they provided a good baseline for us to compare our crowdsourcing-inspired icons against. However, they represent but a single designer's work and therefore are not representative of the entire profession. It is possible that icons from a different designer would yield substantially different results.

Finally, the concepts that we examined consist of use cases for ubiquitous sensing platforms that are likely to be supported in the near term. Because these technologies are still under development and we cannot possibly know what the "killer applications" are going to be ten years from now, it is possible—indeed likely—that many more privacy-sensitive use cases are likely to be identified in the coming years. While our findings are likely to be relevant as these platforms continue to be developed, they are by no means complete. More work needs to be performed to thoroughly understand the range of privacy risks associated with these technologies, how users perceive those risks, and how the relevant information can best be communicated to users.

Conclusion and Future Work

We present a crowdsourced approach to the design of privacy icons for a ubiquitous sensing platform. These icons are meant to communicate specific data collection scenarios to

[3] We cannot perform a direct statistical comparison because different indicators were examined, at different times, on different subjects.

users, for instance, video recording or gender detection. We experimented with a three-stage process where (1) Mechanical Turkers would design icons, (2) synthesis of new icons from the Mechanical Turk icons, and (3) Mechanical Turk evaluation of the synthesized icons. We compare our process with a set of designer-drawn icons, and for each of our scenarios the crowdsourcing approach performed no worse and sometimes significantly better.

In the future, one can imagine a multitude of sensors that are constantly being used to make inferences about and detect objects in the users' environment; some sensors may be used for redundancy or improved accuracy, for instance, simultaneously using both audio and video to identify a user. For access control purposes, all of these sensors will need to be managed by the platform and abstracted away so that applications need only access the resulting data. In this case, the role of a privacy indicator is to convey what information is being accessed by the application, not what sensors were used by the platform to extract that information. We observed that when we examined indicators that included additional information about how the data was extracted (i.e., what sensors were used), it was a distraction that resulted in lower comprehension rates. Participants focused on this information instead of the more important elements—what data was being accessed. Thus, our results suggest that these types of privacy indicators should be designed to convey *what* information will be used by applications, not *how* it will be collected. However, additional studies are needed to examine whether users really care about this distinction.

Another interesting area is the difficulty in disambiguating the face detection and face recognition concepts. Future work needs to be conducted to examine potential users' risk concerns surrounding these data collection concepts.

ACKNOWLEDGEMENTS

We would like to thank Jacob Carter for creating the initial icon set and the Intel Software and Services Group (SSG) for their assistance. This work was supported by the Intel Science and Technology Center for Secure Computing (ISTC-SC).

REFERENCES

1. G. D. Abowd, G. R. Hayes, G. Iachello, J. A. Kientz, S. N. Patel, M. M. Stevens, and K. N. Truong. Prototypes and paratypes: Designing mobile and ubiquitous computing applications. *IEEE Pervasive Computing*, 4(4):67–73, Oct. 2005.
2. G. D. Abowd and E. D. Mynatt. Charting past, present, and future research in ubiquitous computing. *ACM Trans. Comput.-Hum. Interact.*, 7(1):29–58, Mar. 2000.
3. V. Bellotti and A. Sellen. Design for privacy in ubiquitous computing environments. In *Proceedings of the third conference on European Conference on Computer-Supported Cooperative Work*, pages 77–92, Norwell, MA, USA, 1993. Kluwer Academic Publishers.
4. A. R. Beresford, A. Rice, N. Skehin, and R. Sohan. Mockdroid: Trading privacy for application functionality on smartphones. In *Proceedings of the 12th Workshop on Mobile Computing Systems and Applications*, HotMobile '11, pages 49–54, New York, NY, USA, 2011. ACM.
5. M. S. Bernstein, G. Little, R. C. Miller, B. Hartmann, M. S. Ackerman, D. R. Karger, D. Crowell, and K. Panovich. Soylent: A word processor with a crowd inside. In *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology*, UIST '10, pages 313–322, New York, NY, USA, 2010. ACM.
6. M. B. Curry, S. J. McDougall, and O. de Bruijn. The effects of the visual metaphor in determining icon efficacy. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 42, pages 1590–1594. SAGE Publications, 1998.
7. L. D'Antoni, A. Dunn, S. Jana, T. Kohno, B. Livshits, D. Molnar, A. Moshchuk, E. Ofek, F. Roesner, S. Saponas, M. Veanes, and H. J. Wang. Operating system support for augmented reality applications. In *Proceedings of the 14th USENIX Conference on Hot Topics in Operating Systems*, HotOS'13, pages 21–21, Berkeley, CA, USA, 2013. USENIX Association.
8. S. P. Dow, A. Glassco, J. Kass, M. Schwarz, D. L. Schwartz, and S. R. Klemmer. Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. *ACM Trans. Comput.-Hum. Interact.*, 17(4):18:1–18:24, Dec. 2010.
9. W. K. Edwards and R. E. Grinter. At home with ubiquitous computing: Seven challenges. In *Proceedings of the 3rd International Conference on Ubiquitous Computing*, UbiComp '01, pages 256–272, London, UK, UK, 2001. Springer-Verlag.
10. W. Enck, P. Gilbert, B.-G. Chun, L. P. Cox, J. Jung, P. McDaniel, and A. N. Sheth. Taintdroid: an information-flow tracking system for realtime privacy monitoring on smartphones. In *Proceedings of the 9th USENIX conference on Operating systems design and implementation*, OSDI'10, pages 1–6, Berkeley, CA, USA, 2010. USENIX Association.
11. A. P. Felt, E. Ha, S. Egelman, A. Haney, E. Chin, and D. Wagner. Android permissions: user attention, comprehension, and behavior. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, SOUPS '12, New York, NY, USA, 2012. ACM.
12. S. Guha, M. Jain, and V. N. Padmanabhan. Koi: A location-privacy platform for smartphone apps. In *Proceedings of the 9th USENIX Conference on Networked Systems Design and Implementation*, NSDI'12, pages 14–14, Berkeley, CA, USA, 2012. USENIX Association.
13. J. I. Hong, J. D. Ng, S. Lederer, and J. A. Landay. Privacy risk models for designing privacy-sensitive ubiquitous computing systems. In *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, DIS '04, pages 91–100, New York, NY, USA, 2004. ACM.

14. P. Hornyack, S. Han, J. Jung, S. Schechter, and D. Wetherall. These aren't the droids you're looking for: retrofitting android to protect data from imperious applications. In *Proceedings of the 18th ACM conference on Computer and communications security*, CCS '11, pages 639–652, New York, NY, USA, 2011. ACM.

15. G. Iachello and G. D. Abowd. Privacy and proportionality: Adapting legal evaluation techniques to inform design in ubiquitous computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, pages 91–100, New York, NY, USA, 2005. ACM.

16. Intel Corporation. Intel RealSense Technology, 2014. `http://www.intel.com/content/www/us/en/architecture-and-technology/realsense-overview.html`.

17. R. Kang, S. Brown, L. Dabbish, and S. Kiesler. Privacy Attitudes of Mechanical Turk Workers and the U.S. Public. pages 37–49. USENIX Association, Submitted.

18. R. Kannavara and K. Shippy. Topics in Biometric Human-Machine Interaction Security. *Potentials, IEEE*, 32(6):18–25, 2013.

19. P. G. Kelley. Conducting Usable Privacy & Security Studies with Amazon's Mechanical Turk. In *Proceedings of the 6th Symposium on Usable Privacy and Security (SOUPS)*, SOUPS 2010. USENIX Association, 2010.

20. T. Kindberg and A. Fox. System software for ubiquitous computing. *IEEE Pervasive Computing*, 1(1):70–81, Jan. 2002.

21. A. Kittur, E. H. Chi, and B. Suh. Crowdsourcing User Studies with Mechanical Turk. In *CHI '08: Proceeding of The Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems*, pages 453–456, New York, NY, USA, 2008. ACM.

22. R. Kumar, J. O. Talton, S. Ahmad, and S. R. Klemmer. Bricolage: Example-based retargeting for web design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 2197–2206, New York, NY, USA, 2011. ACM.

23. M. Langheinrich. Privacy by design - principles of privacy-aware ubiquitous systems. In *Proceedings of the 3rd international conference on Ubiquitous Computing*, UbiComp '01, pages 273–291, London, UK, UK, 2001. Springer-Verlag.

24. S. Lederer, J. I. Hong, A. K. Dey, and J. A. Landay. Personal privacy through understanding and action: Five pitfalls for designers. *Personal Ubiquitous Comput.*, 8(6):440–454, Nov. 2004.

25. J. Lin, N. Sadeh, S. Amini, J. Lindqvist, J. I. Hong, and J. Zhang. Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, UbiComp '12, pages 501–510, New York, NY, USA, 2012. ACM.

26. Microsoft Corporation. Xbox One, 2014. `http://www.xbox.com/en-US/xbox-one/innovation`.

27. J. Moyes and P. W. Jordan. Icon design and its effect on guessability, learnability, and experienced user performance. *People and computers*, (8):49–60, 1993.

28. J. Newman. Pebble smartwatch pre-orders are sold out, $10+ million pledged. Time, May 10 2012. `http://techland.time.com/2012/05/10/pebble-smartwatch-pre-orders-sold-out/`.

29. L. Palen and P. Dourish. Unpacking "privacy" for a networked world. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM New York, NY, USA, April 5-10 2003.

30. A. Raij, A. Ghosh, S. Kumar, and M. Srivastava. Privacy risks emerging from the adoption of innocuous wearable sensors in the mobile environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 11–20, New York, NY, USA, 2011. ACM.

31. F. Roesner, T. Kohno, and D. Molnar. Security and privacy for augmented reality systems. *Communications of The ACM*, 2014. `http://www.franziroesner.com/pdf/arsec-cacm2014-preprint.pdf`.

32. J. H. Saltzer and M. D. Schroeder. The protection of information in computer systems. *Proceedings of the IEEE*, 63(9):1278–1308, 1975.

33. J. Stern. Google glass explorer edition to ship this month. ABC News, April 10 2013. `http://abcnews.go.com/blogs/technology/2013/04/google-glass-explorer-edition-to-ship-this-month/`.

34. C. Thompson, M. Johnson, S. Egelman, D. Wagner, and J. King. When it's better to ask forgiveness than get permission: Designing usable audit mechanisms for mobile permissions. In *Proceedings of the 2013 Symposium on Usable Privacy and Security (SOUPS)*, 2013.

35. L. von Ahn and L. Dabbish. Labeling images with a computer game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 319–326, New York, NY, USA, 2004. ACM.

36. M. Weiser. The computer for the 21st century. *Scientific American*, pages 94–104, 1991.