A Scalable Inclusive Security Intervention to Center Marginalized & Vulnerable Populations in Security & Privacy Design

Mattea Sim
matsim@iu.edu
Indiana University
Bloomington, IN, USA

Kurt Hugenberg
khugenb@iu.edu
Indiana University
Bloomington, IN, USA

Tadayoshi Kohno
yoshi@cs.washington.edu
University of Washington
Seattle, WA, USA

Franziska Roesner
franzi@cs.washington.edu
University of Washington
Seattle, WA, USA

ABSTRACT

Research in computer security has increasingly considered the needs of marginalized and vulnerable groups in technology. Through this work, we hope to translate this research movement into practice and, ultimately, cause designers-in-training (and, eventually, designers) to consider a more inclusive range of stakeholders. Thus, we created an educational intervention to center marginalized and vulnerable populations in the context of threat modeling. We find that computer security students are more likely to consider unique threats and vulnerabilities facing marginalized and vulnerable populations after being exposed to an intervention prompting them to think about populations that might often be overlooked. We suggest practical methods to teach designers-in-training inclusive methods in computer security and discuss other possible adoptions of this practice across the field. This work is part of an important shift toward inclusive security that centers marginalized and vulnerable populations both in research and in practice.

ACM Reference Format:
Mattea Sim, Kurt Hugenberg, Tadayoshi Kohno, and Franziska Roesner. 2023. A Scalable Inclusive Security Intervention to Center Marginalized & Vulnerable Populations in Security & Privacy Design. In *New Security Paradigms Workshop (NSPW '23), September 18–21, 2023, Segovia, Spain.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3633500.3633508

1 INTRODUCTION

Research in computer security has, since its inception, focused on technical elements of computer security and privacy protections, vulnerabilities, and risks. However, since the introduction of the concept of "usable security" [78, 79], the field of usable security, computer security, and human-computer interaction more broadly have begun to focus on how to build secure systems with a human-centric focus. This movement has evolved to focusing on not just a "user" in the abstract "default" sense, but, given rising interest in inclusive security, security and privacy researchers are beginning to more fully consider marginalized and vulnerable (i.e., M&V) stakeholders [65, 71, 73]. We define marginalized populations as groups of people who are often excluded from mainstream social, economic, and or cultural life. We define vulnerable populations as groups of people who are uniquely susceptible to coercion or attacks and do not often have the socio-cultural power or resources to deal with those attacks.[1] Centering the needs of marginalized and vulnerable users in security and privacy is important for inclusive design because the needs of these populations are often ignored and the threats facing these populations can sometimes be unique.

While this shift has manifested in the computer security and privacy *research* community, in the present work we seek to facilitate a conversation around the following question: how can a focus on marginalized and vulnerable users in computer security and privacy be translated into *practice*? How can we support and empower students, developers, and practitioners to deeply consider specific marginalized and vulnerable stakeholders in their work, rather than thinking consciously or unconsciously only about a "default" persona (i.e., a culturally prototypic user, often straight white tech-savvy men [25, 33, 68])?

This project is a collaboration between two computer security researchers (T.K. and F.R.), who teach an undergraduate computer security course at their institution, and two social psychologists at another institution (K.H. and M.S.), who specialize in studying stereotyping. Teaching threat modeling is central to one of our institution's undergraduate computer security curriculum (University of Washington) and is a key early step of successful secure system design. Thus, across two studies (an initial study and a replication), we investigate whether prompting students to consider overlooked populations increases their likelihood of considering marginalized and vulnerable populations during the threat modeling process. These studies took place in two undergraduate computer security courses (i.e., for designers-in-training). In short, will prompting

NSPW '23, September 18–21, 2023, Segovia, Spain

ACM ISBN 979-8-4007-1620-1/23/09…$15.00
https://doi.org/10.1145/3633500.3633508

[1] We recognize that the term "vulnerable" has different definitions across communities, and is often associated with the susceptibility of computing systems to attacks. In our case, we define vulnerable populations along the axes of both disproportionate computing threat and a lack of socio-cultural power, which includes disproportionately disadvantaged populations that do not fall under the umbrella of the term "marginalized populations." Other definitions of vulnerable populations may also include higher-power groups such as celebrities or government officials.

designers-in-training to consider users who are often left out during the design process lead to a more inclusive set of possible threats and use cases?

While the answer to this question might be an intuitive "yes", without concrete data, we cannot be sure. Moreover, concrete data — whether "yes" or "no" — can provide a foundation for future discussions on how to best assist teams in research and industry with proactively identifying marginalized and vulnerable populations related to the technology under consideration. Although the current intervention is deployed in the classroom, we see this work as part of a broader shift in the field toward adopting new approaches to center marginalized and vulnerable populations, facilitating more inclusive computer security practices. In this way, the present intervention is a starting point, rather than a complete solution, to be built upon and help facilitate broader shifts in the field. Indeed, we explore whether we can prompt students to consider often overlooked populations while threat modeling, however, including these populations in the design process is integral to ensuring designers leave the room truly understanding the unique needs of and harms facing marginalized and vulnerable populations [16].

Having established the overall goals of our study, we now formulate our research questions:

> **RQ1:** Does making marginalized and vulnerable populations salient to students increase the extent to which students identify marginalized and vulnerable populations during a threat modeling exercise?
> **RQ2:** At baseline — without prompting any considerations of marginalized and vulnerable populations — to what extent do students identify marginalized and vulnerable populations while conducting a threat modeling exercise?
> **RQ3:** Which specific stakeholders do students identify, both with and without the salience intervention?

We see the present work as important and timely for multiple reasons. First, to foreshadow our findings, we find that making M&V populations salient *does* improve students' ability to identify marginalized and vulnerable populations while conducting threat modeling exercises. We replicate this result across two studies and across two different academic quarters, with different instructors. Put simply, the intervention is highly effective, is easy to implement with little prior training, and is highly scalable. Second, we see the present work as part of a broader disciplinary conversation about how to integrate deep, non-stereotyped considerations of marginalized and vulnerable populations into computer security in practice, specifically in the context of threat modeling. Our IRB-approved experiment and replication demonstrate that one instantiation of such an intervention works. We look forward to future discussions and work that builds on our initial findings and intervention design here.

2 BACKGROUND AND RELATED WORK

2.1 Default Persona and Stereotypes in Computing Design

Technology design and implementation often centers certain populations — an assumed typical user or "default persona" — while overlooking others. However, the populations included in this default persona often reflect the demographics of higher-power groups (e.g., white, male, upper-class, non-disabled) [9, 14, 67, 68], and similarly of the designers themselves, effectively excluding marginalized and vulnerable populations. Indeed, some social groups are often culturally "invisible" across multiple contexts, with Asian Americans and Black women often failing to "come to mind" or failing to appear in cultural representations because they are not the cultural default [25, 33, 75]. Notably, the present work was situated in the U.S., which affects our assumptions about predominantly U.S.– and Western– centric cultural defaults. Default personas can be specific to a given cultural context, and thus the default persona may look different in other cultures [9, 39].

Computing design is not immune to the effects of this default persona. Scholars have emphasized, for example, how racism and racial discrimination are deeply embedded in algorithmic training data, which can reinforce and reproduce inequality [9], and how historical gender data gaps exclude women from much of both technological and broader societal design [18]. These conversations have gained traction in recent years, with increasingly prominent examples of the default persona in practice and the inevitable discrimination that follows. For instance, face detection technology has frequently been subject to critique for racial and gender biases (e.g., higher gender classification errors for people with darker skin tones [13]). Here, the training data often include predominantly white faces, leaving many racial groups underrepresented and producing systems that simply recognize white faces more effectively. The use of these face detection technologies in the criminal justice system has produced some of the most severe consequences, with Black Americans erroneously jailed due to these biased systems [61]. Recently, because such face recognition systems are used by the U.S. Customs and Border Protection app, this has made it difficult for Black asylum seekers to apply for asylum in the U.S. [20]. As another example, older adults and users with disabilities are often left out of design considerations as well [14, 67]. For instance, social media platforms often fail to ensure content is accessible, including misinformation labels that may be less effective at communicating these important warnings to low vision and blind users [67]. Designing for the default persona and failing to include marginalized and vulnerable groups creates systems that discriminate by design.

Psychological research also provides evidence for this practice of "defaulting." People tend to hold default prototypes for what a "typical" group member looks like, which is often grounded in stereotypes. For instance, people have a stereotypical racial prototype about what the typical "American" looks like, assuming that white Americans are more American than Asian or Black American [25].[2] These default assumptions often occur quickly and unintentionally, as evidenced by methods showing people's automatic associations between groups and their prototypes (e.g., Americans and whiteness). As another example, Black women are often subject to intersectional invisibility, whereby they are seen as prototypical of neither their race nor their gender and are consequently overlooked in a variety of social contexts and in research or civil movements on racism and sexism [17, 63, 66]. Indeed, those

[2]Although "American" could include North and South America, we use the term "American" here to refer to "a person in the United States," consistent with language in cited works.

who do not fit into a group's default prototype are rendered socially invisible. Default prototypes, however, frequently represent higher-status and more privileged groups, leaving marginalized groups invisible. The same is true in technology, in which countless systems, services, and devices have been designed for a higher-status default persona that renders them less effective at best, and dangerous at worst, for marginalized and vulnerable groups.

2.2 Using Salience Interventions to Reduce Biases

Default personas can be particularly insidious because, as noted above, they can come to mind unintentionally and automatically. In our research, we hope to provide tools to instead intentionally consider marginalized and vulnerable populations during a threat modeling exercise. In psychological terms, this intervention is designed to increase the *salience* — the prominence or conspicuousness — of marginalized and vulnerable populations, which past research has reliably shown will increase attention and processing [35].

Research on prejudice reduction interventions, a related but distinct form of intervention from the present work, shows that these intentional practices can be effective in a variety of contexts. Indeed, motivated individuals can reduce unintentional biases by becoming aware of the bias and its consequences, and subsequently implementing intentional strategies to reduce the bias [23]. This type of educational intervention used in other contexts reduced implicit racial bias [23] and increased hiring of women in STEM departments [24]. Thus, bringing attention to the default persona and its consequences, while also providing a strategy to mitigate this tendency to default, may be a useful approach for intervention in computer security education.

Increasing the salience of other groups beyond the default persona may also help computer scientists consider marginalized and vulnerable populations. The degree to which social categories are mentally accessible and seem relevant to the task affect whether these categories will be *salient* to perceivers [12, 47]. A threat modeling task may make the default persona, but not M&V populations, more salient to computer scientists. Indeed, group members seen as less prototypical often go forgotten, unnoticed, and unheard [66]. However, interventions that manipulate the salience of M&V groups, by explaining how technology design often overlooks these groups and causes unique harms, may help computer scientists focus more on M&V populations.

One methodology for proactively identifying populations is through stakeholder analyses, as is central to approaches like Value Sensitive Design [26, 27]. Such approaches encourage deep consideration of how technology impacts a wide range of stakeholders, and whether it supports their human values (e.g., privacy, autonomy, informed consent). Because technology can create disparate harms across groups, the stakeholder groups and impacts that come to computer scientists' minds are particularly appropriate points at which to intervene on the default persona and encourage deeper consideration of how technology impacts M&V stakeholders.

2.3 Computer Security Education and Toolkits

There is also an active field of research and practice around computer security education for students and toolkits for practitioners.

The literature on security education is vast, including dedicated publication venues, such as the World Conference on Information Security Education [11], along with regular publications of security-related educational material at broad CS-education conferences, such as ACM Special Interest Group in Computer Science Education (SIGCSE). For example, specific interventions or techniques that have been studied include interventions around ethics [62], narrative storytelling [45], science fiction prototyping for considering broader societal issues surrounding computer systems [41], and using Capture the Flag, board games, and hands-on exercises [54]. Our past work has also explored fiction as a vehicle for surfacing the harms of designing for a default persona [39, 40]. While our present work was done in the classroom and has implications for future classroom instruction, the focus of our work is not on classroom education but what we as a field can learn about the role of interventions in leading designers to proactively consider M&V populations.

On the toolkit side, researchers and computer security professional have developed threat modeling tools and toolkits. Example tools and toolkits include STRIDE [38], Persona non Grata [15], and the Security Cards [21], as well as a hybrid approach that combines different tools and toolkits [53].

To our knowledge, none of these prior educational or toolkit efforts have directly targeted the inclusion of M&V populations in threat modeling; we aim to bridge that gap here.

2.4 Computer Security and Privacy for M&V Populations

Although educational resources have not always centered underserved populations, recent research efforts focus on just this issue. In recent years, a growing body of work has begun to foreground the study of the needs of marginalized and vulnerable populations. For example, in no particular order and not exhaustively, researchers have studied user groups such as older adults [28, 36, 51, 56], children [19, 30, 37, 42, 43, 50, 52, 55, 76], sex workers [6, 34, 49], people with visual impairments [1–4], survivors of domestic abuse [7, 72, 77], undocumented immigrants [32], refugees [69], queer people [29], transgender people [44], and incarcerated people [58] and people under electronic monitoring [57]. Early systematizations of this space can be found in [65, 71, 73]. Several workshops related to this topic have sprung up, including the Workshop on Inclusive Privacy and Security (WIPS) and the Workshop on Security for Harassment Online, Protections, and Empowerment (SecHOPE). We are excited about all of the important work in this area of security and privacy for marginalized and vulnerable people, and with this paper, we aim to contribute to conversations about how to operationalize this perspective beyond research.

3 METHODOLOGY

3.1 Participants

Participants (Study 1 N = 117, Study 2 N = 108) were computer science undergraduates enrolled in either the Autumn 2022 (Study 1) or Winter 2023 (Study 2) quarter of an upper-level computer security course (i.e., typically undergraduates in their third or fourth year) taught by the two primary investigators (one investigator one quarter, the other investigator the other quarter). Each course had

four lab sections led by teaching assistants. Students completed the study as part of an in-lab threat modeling exercise. Students were asked to optionally provide demographic information at the end of the quarter, however, less than 10% of students did so and it is therefore not discussed further.[3]

3.2 Ethics

The study was reviewed and approved by the university's Human Subjects Review Board (IRB). Responses were anonymized before the data were analyzed, and students were given the opportunity to opt out of having their data included in the study (1 student in the Autumn and 0 students in the Winter opted out). See Appendix for the email sent to students allowing them to opt-out. Opt-outs were not requested and data were not analyzed until after grades had been finalized, and it was made clear to students that their decision to opt out would have no impact on their grade in the course. Because the lab sections assigned to the control condition were assigned to the intervention condition later in the quarter (and vice versa), all students had the opportunity to be exposed to all educational materials if they attended both classes.

3.3 Procedure

Our latest materials, including this intervention, can be found at https://security-education.cs.washington.edu/.

Two studies were conducted with identical procedures, allowing us to investigate if our results would replicate with a new set of students and different instructors.

The procedure involved an in-lab threat modeling assignment typical in computer security courses. This assignment was an ungraded activity. Here, students were asked to focus on an augmented reality (i.e., AR) headset and to consider the security and privacy concerns of this technology. We chose this technology because it is currently emerging and presents many potential security and privacy and safety risks, and because our lab has existing expertise in this space [64]. Students responded to several threat modeling questions to identify the security goals of the headset, the assets to be protected, the adversaries who might attack this headset and their goals, and potential threats or vulnerabilities. Teaching assistants prompted the students to complete the assignment on Canvas, such that all materials were administered via the online assignment. Each student answered the threat modeling questions (in which we coded our primary dependent variable) independently via the assignment submitted online. Students also self-selected other students to discuss responses, but all questions in the assignment assessed here were completed outside of the group, by each student individually.

We employed a mixed-model design, with both within-subjects and between-subjects components. During each class' lab sections during the first week of the quarter, students completed an online threat modeling assignment twice (Time 1 vs. Time 2, a within-subjects variable). At Time 1, all students submitted answers for the AR headset threat modeling exercise as described above (see appendix for full materials) without additional prompting, serving as a baseline for the extent to which students spontaneously consider marginalized and vulnerable populations while threat modeling.

Immediately after this, at Time 2, participants completed the same assignment online again. However, we manipulated on a between-subjects basis whether students were in the Control or the Salience Intervention condition. Participants in the Control condition (Study 1 n = 64, Study 2 n = 45) read a prompt on Canvas explaining that sometimes when people consider a question a second time, they come up with different responses. Students considered the questions a second time and submitted new answers. Participants in the Salience Intervention condition (Study 1 n = 53, Study 2 n = 63) read an educational prompt on Canvas that explained the default persona and asked students to consider populations that might often be overlooked. In more detail, the intervention prompt explained that designers often unintentionally design for some populations and not others, and provided three examples of this default persona in practice (i.e., crash test dummies matching the anatomy of adult males while excluding much of the population; face recognition technologies without gender or racial diversity; smartphones designed for a single user and excluding parent-children sharing, lower-income contexts, or non-U.S. contexts). Intervention condition students were prompted to "think about populations that engineers might not normally think about" while submitting answers to the same threat modeling questions a second time.[4]

Each lab section was randomly assigned to the Control or Salience Intervention conditions. The study was conducted with no primary investigators present in the classroom, and teaching assistants who ran the lab sections were blind to hypotheses.

Expert Panel. Separate from the in-lab activity with student designers-in-training, an expert panel completed the same threat modeling exercise, albeit with a different procedure. The purpose of the panel was to generate broader stakeholder categories and themes by which to later understand students' responses. The expert panel exercise consisted of three principal investigators and six experts in computer security, including experts in both M&V populations and AR. The team generated a list of stakeholders, use cases, assets, adversaries, and threats for AR headsets, with 5–10 minutes of individual brainstorming per category before discussing together as a group. Experts wrote ideas on sticky notes and added them to a collaborative brainstorming wall (see Figure 1). The team was also asked to think expansively about stakeholders, including M&V populations. Following the expert panel exercise, the primary investigators clustered experts' responses into 18 distinct categories of stakeholders. An additional "Other" category was created for responses that did not fall within existing categories (see Table 2 in Appendix C for all categories and examples listed by experts;

[3]Although we do not have demographics for students in these studies, demographic data are available for undergraduates in the Paul G. Allen School of Computer Science and Engineering at UW [60]. In the beginning of the academic year in which these studies were conducted, the students were predominantly male (66%, and 34% female) and classified as non-underrepresented minority populations (78%). 10% of students were classified as underrepresented minority populations (i.e., African American, American Indian/Alaska Native, Hawaiian/Pacific Islander and Latinx/Hispanic), 12% of students were international, and 27% were first-generation college students.

[4]Students had the opportunity to complete a similar threat modeling assignment later in the quarter where their class section was assigned to the opposite condition (i.e., students who were initially in the Control condition completed the Salience Intervention condition), however, student drop-off was quite high due to lowered class attendance, such that there were not enough students present to include these data in analyses.

Figure 1: Photo of the expert panel exercise, showing sticky notes of stakeholders produced by the experts.

notably, some students' responses also included stakeholders that clearly fell under an existing category, but were not always listed as an example by the experts). Thus, the expert panel generated 18 categories, for a total of 19 categories including the "Other" category.

Importantly, the stakeholder list generated by the expert panel was not intended as an exhaustive or complete list of all possible stakeholders, but rather as a tool by which we could categorize and understand students' responses. The resulting list has gaps and would likely look different if a new panel of experts completed the same task. Thus, our stakeholder list is not intended as itself a primary contribution of this work, although the process of expert threat modeling may be a useful component of the overall toolkit procedure.

3.4 Analyses

We coded students' responses to the threat modeling exercise along several dimensions. Of primary interest was understanding the proportion of all stakeholders identified by students that belonged to marginalized and vulnerable populations. To calculate this, an investigator coded each student's response by the total number of stakeholders they identified and the number of M&V stakeholders they identified. These were used to calculate a proportion score (number of M&V stakeholders / total number of stakeholders identified) for each student that served as the primary dependent variable in a mixed-model ANOVA. Proportion variables such as this are common practice in psychological research [22, 46]. For descriptive analyses, we also created a variable indicating whether each student discussed at least one M&V stakeholder or no M&V stakeholders.

Of secondary interest was understanding the broader categories of stakeholders that students identified in the exercise to answer RQ3. For this analysis, we relied on the clusters of categories generated by the expert panel. Thus, an investigator also coded each response by listing all stakeholders identified in the response, and subsequently categorizing each identified stakeholder into one of the 19 broader stakeholder categories.

Because students were not directly asked to identify stakeholders, the identification of stakeholders was sometimes implied in responses through identification of adversaries (e.g., an abusive partner implies a victim/survivor of abuse as a stakeholder) or adversary actions (e.g., stalking implies a victim of stalking) or assets (e.g., protecting game data implies gamers as a stakeholder). Thus, when discussing marginalized and vulnerable stakeholders identified by students, we are broadly considering both directly identified stakeholders and identified adversaries, actions, or assets that uniquely affect marginalized and vulnerable stakeholders.

4 RESULTS

4.1 RQ1: Students consider M&V stakeholders more after a salience intervention

Our primary research question was whether students consider marginalized and vulnerable stakeholders more after a salience intervention. To investigate this question, we first conducted a 2 (Response Time: Time 1, Time 2) x 2 (Condition: Intervention, Control) mixed-model ANOVA on the proportion of stakeholders identified belonging to marginalized and vulnerable (M&V) groups. In Study 1, there were significant main effects of Response Time, $F(1, 105) = 39.069$, $p < .001$, $n_p^2 = .271$, and Condition, $F(1, 105) = 19.547$, $p < .001$, $n_p^2 = .157$, qualified by a significant interaction between Response Time and Condition, $F(1, 105) = 41.460$, $p < .001$, $n_p^2 = .283$. We replicated this interaction between Time and Condition in Study 2, $F(1, 97) = 19.453$, $p < .001$, $n_p^2 = .167$ (see Figure 2). In this and subsequent sections, we decompose the interactions with t-tests and descriptive analyses to answer each research question of interest.

To understand whether students considered M&V stakeholders more after the salience intervention, we tested how students' responses at Time 2 differed depending on if they read a control prompt or the intervention prompt. In Study 1, the proportion of stakeholders identified belonging to M&V groups was significantly higher amongst students in the Intervention condition ($M = .42$, $SD = .42$) as compared to those in the Control condition ($M = .07$, $SD = .16$), $t(105) = 6.016$, $p < .001$, 95% CI [0.76, 1.58], $d = 1.17$. After reading the intervention prompt, 50.9% of students identified at least one M&V stakeholder, as opposed to 15.6% of students who did so in the Control condition. Results from Study 2 replicated this pattern. The proportion of stakeholders belonging to M&V groups was again significantly higher after the intervention ($M = .25$, $SD = .35$) compared to the control prompt ($M = .01$, $SD = .05$), $t(97) = 4.403$, $p < .001$, 95% CI [0.48, 1.32], $d = 0.90$. Here, 36.5% of students after intervention identified at least one M&V stakeholder, as opposed to just 2.2% of students who did so after the control prompt. These results suggest a salience intervention holds promise for helping students consider M&V populations whom they might initially overlook. Directly prompting students to consider a more diverse range of stakeholders increases the degree to which they identify marginalized and vulnerable stakeholders and unique threats facing marginalized and vulnerable stakeholders.

We conducted additional analyses to further investigate RQ1 and test the effectiveness of the intervention. We conducted within-subjects analyses to capture students' changes in responses across

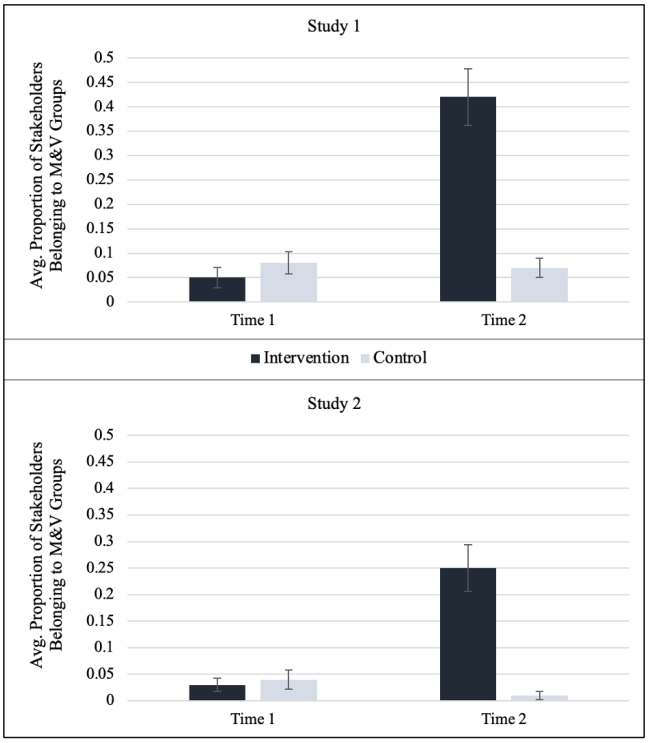

Figure 2: Interaction between response time and condition on the average proportion of stakeholders identified belonging to M&V groups in Study 1 (top panel) and Study 2 (bottom panel). Error bars represented by standard error of the mean.

Time 1 and Time 2, allowing us to investigate whether the same students would be more likely to identify M&V stakeholders after an intervention as compared to their previously unprompted responses to the same questions. In Study 1, students identified a significantly higher proportion of stakeholders belonging to M&V groups after reading the intervention prompt at Time 2 as opposed to the regular prompt at Time 1, $t(46) = -6.002$, $p < .001$, 95% CI [-1.21, -0.54], $d = -0.88$. This same pattern was replicated in Study 2, $t(57) = -4.835$, $p < .001$, 95% CI [-0.92, -0.35], $d = -0.64$. In contrast, amongst students in the Control condition, there was no significant difference in this proportion when comparing between Time 1 and Time 2, Study 1: $t(59) = 0.299$, $p = .766$, 95% CI [-0.22, 0.29], $d = 0.04$; Study 2: $t(40) = 1.482$, $p = .146$, 95% CI [-0.08, 0.54], $d = 0.23$. In other words, merely being prompted to answer the threat modeling questions a second time was not enough to elicit greater consideration of M&V stakeholders. Instead, a prompt that explicates the importance of considering underserved populations does, in fact, help students consider marginalized and vulnerable populations to a greater degree while threat modeling. Further, whereas students may default to considering an abstract (and perhaps implicitly more privileged) user, this intervention empowered students to consider previously overlooked marginalized and vulnerable stakeholders.

Consider the following case study, which provides an example of a student in the Intervention condition actively recalibrating their response after reading the salience prompt. At Time 1, they discussed threats and vulnerabilities facing the default user.

> **Case Study 1 (Time 1):** "The security goals of the AR headset are reliability and usability in addition to protecting privacy and safety. The assets that must be protected are the person wearing the headset, the personal information the user entered into the system, and the safety of the person wearing the headset. Adversaries might be attackers who are trying to steal personal information or data related to the user. Their goal might be to simply steal data or even try to harm the user of the product. Some threats include the camera feature of the product. Attackers can easily find a vulnerability in the software and obtain the camera footage. Another threat could be the attacker altering what the user sees to spread misinformation."

After the salience prompt at Time 2, they more deeply considered how people with disabilities might face unique threats and vulnerabilities.

> **Case Study 1 (Time 2):** "To make the product more secure and usable, a goal should be trying to make it accessible to persons with disabilities. The assets are the personal information. For example, some people might be using the product for fun, but others might be using it to assist them in everyday life so it may contain highly sensitive information. This data must be protected. Adversaries could include people trying to commit hate crimes in addition to trying to simply steal data. AR could be highly dangerous if there are too many vulnerabilities."

This case study provides qualitative evidence for the effectiveness of the salience intervention, illustrating how students were able to reconsider their notion of the typical user to instead consider possible needs or vulnerabilities facing marginalized and vulnerable stakeholders.

4.2 RQ2: Students were unlikely to spontaneously consider M&V stakeholders

Of secondary interest was how frequently students considered marginalized and vulnerable populations spontaneously, *without* the salience intervention. To answer this question, we descriptively investigate responses at Time 1, wherein students across both conditions had not been prompted to consider marginalized and vulnerable populations. We expected that students may tend to focus on the more chronically salient default persona, and thus may be fairly unlikely to consider M&V populations at Time 1. Indeed, across both Studies 1 and 2, only a small percentage of students identified marginalized or vulnerable populations at Time 1. In Study 1, only 14.5% of students identified at least one marginalized or vulnerable stakeholder, whereas 84.6% of students did not identify any marginalized or vulnerable stakeholders. Similarly in Study 2, only 7.4% of students identified at least one marginalized or vulnerable stakeholder, compared to 92.6% of students who identified none. Considering the data another way, the proportion of stakeholders from M&V populations was also fairly low across both conditions

at Time 1. Across both studies, on average less than 10% of stakeholders identified at Time 1 (and Time 2, in the Control condition) belonged to M&V populations, regardless of the assigned condition (see Figure 2). Overall, these findings confirm that students were fairly unlikely to spontaneously consider the needs of M&V stakeholders in a threat modeling exercise.

4.3 RQ3: Exploring who comes to mind with and without the salience intervention

Finally, we sought to explore what stakeholders students identified both with and without the intervention. As we can see in Table 1, two primary patterns emerged, each of which we explore in a separate subsection below. First, the default user was the most commonly identified stakeholder (although this was attenuated in the Salience Intervention condition). Second, the marginalized and vulnerable stakeholders discussed by students tended to cluster in specific ways, and the salience intervention itself made some marginalized and vulnerable users much more likely to be mentioned.

4.3.1 The default user was the most commonly identified stakeholder. Students across conditions most frequently discussed default users in the abstract sense, without identifying more specific user demographics or implicating unique needs or threats that would fall outside of the default persona (not including other stakeholders they may have also identified). Indeed, amongst all responses in the Control condition, 95.3% of responses in Study 1 and 88.9% of responses in Study 2 discussed default users. We see similar patterns for students in the Intervention condition at Time 1, before they were prompted with the intervention. Here, 98.1% of students in Study 1 and 100% of students in Study 2 discussed default users in their threat model.

The following case study segment from Study 2 illustrates how students often discussed the user in an abstract, broad sense.

> **Case Study 2 (Time 1):** "The security goals of the AR headset could be protecting the user's private information (e.g. username, account information, banking information, user activity, etc) as well as audio or video feed of a user's private surroundings and location."

Many responses followed a similar format, detailing vulnerabilities and assets for the "user," without identifying more specific information about who the user is and how their identity might affect the threat model.

Interestingly, students were considerably less likely to discuss a default user after being exposed to the intervention at Time 2, although it still remained the most frequently discussed stakeholder category. Amongst students in the Intervention condition at Time 2, 54.7% of students in Study 1 and 66.7% of students in Study 2 discussed default users. This decreased tendency to focus on a default user might suggest that students are identifying more concrete stakeholder groups following the intervention, even if these stakeholders do not always belong to marginalized and vulnerable populations. Indeed, students prompted with the intervention identified more stakeholder categories overall than did unprompted students, suggesting the intervention might also help students think more expansively during threat modeling, an interesting question for future research.

There were several other categories of stakeholders that students frequently discussed. Without intervention, AR developers/designers (e.g., the AR device company) tended to be the second most frequently discussed stakeholder category. Many students also discussed specific app users (e.g., gamers, drivers using navigation apps), employers and companies (e.g., other companies using AR that might be surveilled), bystanders, and government entities (e.g., military, politicians) as prominent stakeholders. Although there were subtle differences across Studies 1 and 2, students tended to discuss many of the same stakeholder categories.

4.3.2 Students identified specific clusters of marginalized and vulnerable groups. Of additional interest was understanding what marginalized and vulnerable populations students were most likely to identify, both before and after intervention. Interestingly, students tended to identify a specific cluster of marginalized and vulnerable stakeholders, most frequently discussing people with medical or sensory impairments (falling under the category of people with stigmatized social identities [31]; e.g., people with epilepsy, people with visual impairments, people with hearing impairments), vulnerable age groups/people unable to consent (e.g., children), and targets of hate, harassment, and abuse (e.g., victims of abuse or stalking). The categories of M&V stakeholders that students were most likely to identify often differed before and after intervention. For instance, students in Study 1 without intervention were most likely to identify targets of hate, harassment, and abuse (most commonly, victims of abuse or stalking), whereas after the intervention students were more likely to identify vulnerable age groups/people unable to consent (especially children) and people with stigmatized social identities (especially people with disabilities). Students in Study 2 were overall less likely to discuss marginalized and vulnerable categories without intervention, but here again students exposed to the intervention were most likely to identify vulnerable age groups/people unable to consent and people with stigmatized social identities. Consider Case Study 3 below that identifies children as a stakeholder.

> **Case Study 3 (Time 2):** "Once again, the asset of physical safety of the consumer must be protected, but particularly in terms of small children. There might have to be some extra precautions taken in terms of potentially limiting distractions that come up on the headset. Kidnappers could target children who are distracted by something, and children who are actively using an AR headset are definitely distracted and have their line of sight obscured. The vulnerability is taking attention of the user away from the physical world, where there might be certain threats."

At Time 1, Case Study 3 discussed physical safety as an asset for the default user, whereas after exposure to the salience intervention they continued to discuss physical safety as an asset, but now considering how a vulnerable population may have unique considerations for physical safety in AR.

Table 1: Percent of participants within condition who listed each stakeholder category (Note: C=Control, I=Intervention, T=Time).

	Study 1				Study 2			
	CT1	CT2	IT1	IT2	CT1	CT2	IT1	IT2
Default	98%	92%	98%	55%	98%	80%	100%	67%
AR developers/designers	23%	19%	13%	6%	18%	27%	21%	11%
Specific app users	17%	13%	8%	0%	7%	4%	8%	2%
Bystanders	13%	5%	4%	6%	3%	7%	5%	3%
Targets of hate, harassment, and abuse	11%	9%	6%	9%	9%	0%	2%	3%
Government entities	8%	8%	8%	2%	0%	7%	11%	8%
People with stigmatized social identities	5%	6%	4%	19%	0%	2%	3%	19%
Employees/patients	3%	2%	4%	4%	2%	0%	2%	8%
Vulnerable age groups/people unable to consent	2%	0%	2%	34%	0%	0%	2%	14%
Activists/politically involved citizens	2%	3%	0%	0%	0%	0%	0%	0%
People without access to the technology	0%	0%	0%	2%	0%	0%	0%	5%
Non-U.S. citizens	0%	0%	0%	2%	0%	0%	0%	2%
Celebrity and social accounts	0%	0%	2%	0%	0%	2%	0%	0%
Employers and companies	0%	2%	8%	2%	7%	4%	3%	5%
Entities monitoring others	0%	0%	0%	2%	0%	0%	0%	0%
AR 3rd party entities	0%	2%	4%	0%	2%	0%	0%	0%
Other/Uncategorized	0%	0%	0%	6%	0%	0%	2%	6%
AR regulators	0%	0%	0%	0%	0%	0%	0%	0%
Vulnerable workers	0%	0%	0%	0%	0%	0%	0%	0%

5 DISCUSSION

Computing security and privacy is increasingly centering the needs of marginalized and vulnerable communities in design and practice [65, 71, 73]. Across two studies, we investigated how we can use basic tools from social psychology — making salient the needs of marginalized populations [12, 47] — to bring the needs of commonly ignored communities to the forefront of students' minds during a threat modeling exercise. In both studies, advanced computer science undergraduate students completed a threat modeling exercise twice, both before and after a prompt. We manipulated between-participants whether this prompt was a control prompt (simply having students repeat the exercise) or a *M&V salience prompt*, designed to focus students on the needs of M&V populations. We reliably showed that students were much more more likely to consider specific needs and vulnerabilities of M&V groups after the intervention salience prompt 1) as compared to before the prompt or 2) as compared to students who did not receive the prompt. Put simply, at baseline, students are likely to think in terms of the "default persona" (i.e., culturally dominant groups) and not consider the unique needs of or threats faced by M&V populations. However, making M&V populations salient led students to be more likely to center their unique needs and consider how technology uniquely impacts these populations during a threat modeling exercise.

We see the present work as fitting well in the recent wave of research in computing security and privacy that centers the needs of marginalized and vulnerable populations. Here, we find that designers-in-training can be made to think more inclusively by a straightforward prompt. We see this as an important step in moving the recent advances in conceptual knowledge about M&V populations in computing security and privacy [65, 73] into practice with designers-in-training. To this end, we also see the present work as highly *scalable* across classes and highly *deployable* across contexts. The exercise was designed to take less than 10 minutes to complete, it was fully self-contained (i.e., deployed with minimal instruction by teaching assistants other than the primary investigators), and could be done solo or in groups. Further, although the present work was designed with students in mind, this simple intervention could also be used with practicing programmers and designers. Indeed, in our expert panel, we found that having a large number of designers working in a group quickly created a large body of viable considerations for M&V populations' needs and threats using a similar threat modeling exercise. If anything, our trained designers created more and more elaborated considerations of threats to M&V populations. Although the experts' considerations still have gaps, prompting both experts and designers-in-training alike to consider threats and vulnerabilities facing marginalized and vulnerable stakeholders can lead to a more robust treatment of the needs of those populations.

Limitations and Open Questions. We see the present work as part of a continuing conversation about how to center the needs of M&V populations in computing design. That a simple, straightforward intervention can powerfully influence designers-in-training to consider the needs of marginalized groups is an important demonstration in its own right, but there is still extensive research to be done to help establish an agenda for our field. Indeed, we see the present intervention as a starting point, for which there are limitations that can be addressed and built upon in future iterations of this intervention.

First, it is unclear how long the intervention making M&V groups salient is effective. Here, we measured students' responses immediately after the intervention itself. To what extent would the effects

sustain over time? As yet, this is unclear. Other research on diversity training shows that its positive effects often wane over time [59]. It is also clear from work in psychology on concept activation that concepts can be made both temporarily active and chronically active [5]. Whereas some people may not often think of M&V populations' needs, others may do so as a matter of course. Further, the former can become the latter with practice. Indeed, part of becoming a domain expert is forming "habits of mind" that make one think like a domain expert. We argue that consistently centering the needs of M&V populations during design could become part of one's standard practice or of a school's training model in a way that could well make this a chronic habit-of-mind for designers. Future work would benefit from understanding how often such exercises would need to occur, and who would benefit the most from such exercises, to make M&V users come frequently to mind.

In addition, our studies were limited to a carefully controlled intervention in a classroom setting. We hope that the current toolkit can become a part of regular training that may help future designers spontaneously consider M&V populations by habit, a point which we discuss more in the next section. However, there may be several challenges to implementing this intervention beyond the classroom. First, students completed this intervention as part of an in-class assignment. Although the activity was ungraded, classroom dynamics could provide external motivation to satisfy the prompt, rather than the intervention eliciting internal motivation to consider the needs of M&V groups. In industry settings, there will be novel challenges to motivating practicing designers to take part in the intervention without such natural incentives. Future work should also focus on fostering internal motivation to center M&V populations in computing (or assessing whether the current intervention does so naturally) and recruitment methods that account for novel contexts and incentives. Second, the timing of the intervention may also be impactful. Students completed the assignment immediately after reading the salience intervention prompt, allowing the intervention to have its maximum impact on responses. In less controlled settings, less immediacy could mean diminished returns. Future iterations of this intervention should aim to provide designers with the motivation and tools to intentionally intervene on themselves in the future, an effective strategy in other interventions [23, 74]. Third, whereas our intervention helps make M&V groups more salient to *individuals*, it does not account for broader *organizational* goals and priorities. Whether or not an industry's goals align with or diverge from this intervention may affect both its practicality for designers and its likelihood of being adopted by an organization. Thus, it will be integral to focus efforts on both individuals' tendency to consider M&V populations in computing and broader organizational structures that foster or hinder these goals.

We were also unable to measure the demographics of the students and the teaching assistants present during the intervention. The staff demographics were uncontrolled, and with our present data we cannot investigate how this may have affected students' responses. However, our results replicated across two separate quarters and different teaching assistants, suggesting our effects were robust beyond the impact individual teaching assistants may have had on students. Further, it may be the case that students' and designers' own social identities and lived experiences affect the degree to which they spontaneously consider the needs of marginalized and vulnerable populations while threat modeling. Perhaps designers from marginalized and vulnerable populations, or designers with more social contact with these populations, are also more likely to consider often-overlooked M&V populations. These questions will be of considerable interest in understanding who is most likely, and when they are most likely, to consider a diverse range of stakeholders.

Specific features of the intervention are important to acknowledge and consider as well. For instance, we did not directly ask students what stakeholders they believe are impacted by the technology, a question which could have prompted students to consider more specific stakeholder groups. We would suggest that future iterations of this toolkit ask more directly about stakeholders. However, it is unlikely that this question would have prompted students to spontaneously consider M&V populations, who do not appear salient in most students' minds at baseline. In addition, an intervention prompt that is not carefully designed could cause negative reactions, although we did not observe backlash across our two studies. Other staff and students present may be impactful as well. Across both our students and expert panel, a different group composition would likely generate a list of stakeholders distinct from what we presented here. This speaks to the importance of having a diverse group of students, staff, experts, designers, and curriculum when deeply considering the unique harms facing M&V populations. As noted, the current work does not provide a complete list of all possible stakeholders affected by the technology at hand. However, we show that a brief intervention making M&V populations salient leads people to include more M&V populations for consideration in a threat model.

Relatedly, the chosen technology for a threat model will likely impact both the stakeholders that come to mind and people's ability to deeply elaborate on the threat model. For instance, it is possible that AR headsets make certain M&V groups (e.g., children, people with disabilities) especially salient, an effect which may be different when threat modeling for other technologies. Further, students with greater prior experience with AR technologies may be able to more deeply elaborate on different stakeholders and threats pertinent to this technology. Future work should explore the boundaries of these effects across different technologies.

It is also clear that no one designer can consider the needs of all possible users in all possible situations. Thus, although the present work was successful at moving users away from the "default persona" during their threat modeling, it is clear that even amongst M&V populations who come only infrequently to mind, some M&V populations are more salient than others. Specifically, our intervention made vulnerable age groups and stigmatized identities especially salient. This may be due to the specific wording of our intervention, or alternately because concern with protecting others may make children (the most commonly discussed vulnerable age group) spontaneously salient [48]. Future research would benefit from understanding how framing M&V salience differently may make the needs of different groups come to mind.

Interestingly, we observed that after the intervention, students also surfaced a broader set of stakeholders, including non-M&V stakeholders, than otherwise. Though our current study design

does not allow us to investigate it, this observation raises the following research question: Does prompting students to think about marginalized and vulnerable populations cause the students to think more broadly about security and privacy threats, risks, and adversaries, even beyond marginalized and vulnerable populations? That is, does it expand the scope of their threat modeling *in general*?

Finally, our study assessed the frequency with which students identified M&V populations while threat modeling, rather than the extent to which they did so *meaningfully* or *accurately*. For instance, although the present research has been successful at making M&V groups salient to designers, it is not clear that the specific needs that designers were considering were the actual needs of the groups of interest. Put simply, this intervention was successful at making designers consider others' needs, but it is not clear that it was successful at *meeting* others' needs. To do so in a way that does not devolve to a stereotyped treatment of groups' needs, sustained work with populations of interest is needed [10, 16]. Thus, we see the present work not as a panacea, but as a means of making salient to designers that their designs often fall short when M&V populations are not considered. When we seek to actually address that shortfall, we are most likely to succeed when we bring members of M&V populations into the design process. Engaging M&V communities in the threat modeling process may be a starting point to understanding different group's needs for design. Indeed, without including M&V populations in the design process, designers may themselves simply rely on their beliefs about the populations of interest (i.e., stereotypes).

6 FOUNDATIONS AND DIRECTIONS FOR A NEW APPROACH

This work calls for a new approach to threat modeling, one that centers M&V populations. Future iterations of this toolkit could be deployed in a variety of contexts and to a variety of audiences, with promise for extending the impact of this work. In the previous section, we discussed limitations of the current studies and open questions generated by our findings. In the following paragraphs, we highlight several of these open questions to discuss concrete, specific directions and questions posed to the security research community.

First, prompting designers-in-training to intentionally consider overlooked populations may be a promising new approach for teaching threat modeling and other computer design and security practices. Indeed, whereas much of threat modeling focuses on adversaries and threats, threat modeling practices may benefit from deeper consideration of how technology may disparately impact a wide range of stakeholder groups [26, 27]. Our prompt included a description of the default persona and specific examples of how the default excludes other groups, with consequences for design. This direct educational prompt may be useful for designers-in-training to become aware of the defaulting bias and its consequences, a process which prior work has shown can help motivated individuals reduce implicit bias [23]. For this reason, *education* about the default persona and *sustained, prompted practice* considering other stakeholders may be important new tools to integrate into regular curriculum for designers-in-training. Repeated education of this process may help designers-in-training create lasting habits that continue to impact their work beyond the classroom.

Second, as discussed in the previous section, part of adopting this new approach to centering M&V populations in practice includes deploying versions of this intervention outside of the classroom. Designers and other computer scientists in the field may show a similar tendency to spontaneously consider the default persona. This raises the question, how can we engage practicing computer scientists with these exercises? There may be several possible approaches to adopt as a field. For instance, designated conference workshops could include both research centering M&V populations and active participation in prompted exercises to consider how technology uniquely impacts M&V stakeholders. Similar workshops could be created in collaboration with organizations and industry, for adoption in employees' regular training practices. Of considerable interest is discussing how we can move toward intentional practices to proactively center M&V populations in both educational and non-educational settings.

Third, our work and discussion above also raises questions about how to prompt designers to consider M&V populations in a way that appropriately meets their needs, but does not lead to *stereotypical* considerations of their needs. In the short term, we suggest that the first goal of our community is to increase awareness of diverse M&V populations in the design of security technologies and in the threat modeling process; this paper is a step in that direction. Drawing from other disciplines, e.g., Design Justice, after M&V populations are identified, a current best practice for mitigating harms from a reliance on stereotypes is to involve members of those M&V populations directly in the design and evaluation considerations [16]. For example, researchers and practitioners could involve members of M&V populations in the threat modeling process [70]. At NSPW 2023 alone, researchers interviewed low vision and blind users about their experiences with misinformation labels on social media, surfacing a diverse range of both accessibility concerns and proposals for design solutions [67]. Other researchers at NSPW discussed inclusion of M&V populations through the lens of basic capabilities, or identifying basic security hygiene behaviors as fundamental human rights and working with M&V populations to identify unique barriers to these capabilities [14]. There are a diversity of approaches for involving at-risk users in security and privacy research, and we encourage researchers (and practitioners) to consider the full spectrum of options [8].

Finally, what methods can help us evaluate the success of these interventions? Our analysis focused on designers-in-training for the duration of a single class, which leads to the question: how can we know if these interventions impact real production designs? Longitudinal approaches to both the deployment of interventions and the measurement of their impact across time and contexts may be a promising avenue. For instance, students' responses to later assignments could be assessed to understand if the brief intervention had a lasting impact on their tendency to consider M&V populations. Although such methods can be difficult in practice, centering M&V populations as part of regular exercises from designers-in-training to working programmers may lead to measurable changes in production. We hope to discuss methods for the field to center M&V populations in design across development and time course.

7 CONCLUSION

Across two studies, we find that making marginalized and vulnerable populations salient during a threat modeling exercise substantially increases the likelihood of designers-in-training considering these populations. This work lays a foundation for future explorations of how best to intervene in the design process to center the needs of M&V populations. We believe that the present work is important to help set the agenda for security and privacy research to translate good theory into scalable, affordable, and effective practice. Expanding our educational tools to incorporate inclusive stakeholder analyses may be a promising new approach for the field to center marginalized and vulnerable populations in practice.

ACKNOWLEDGMENTS

This work was done as part of the Center for Privacy and Security for Marginalized and Vulnerable Populations (PRISM), supported by the National Science Foundation under Awards SaTC-2205171 and 2207019. We thank Tristan Caulfield and Partha Das Chowdhury for shepherding this paper and the anonymous reviewers for their insightful feedback. We thank the NSPW 2023 attendees for the insightful discussions. We thank Kaiming Cheng, Inyoung Cheong, Rachel McAmis, Kentrell Owens, and Miranda Wei for their participation in the expert panel. We thank the TAs, Aroosh Kumar, Tim Mandzyuk, Kentrell Owens, Noah Ponto, Basia Radka, John Taggart, and William Travis for facilitating the in-class assignment.

REFERENCES

[1] T. Ahmed, R. Hoyle, P. Shaffer, D. Connelly, K. Crandall, and A. Kapadia. Understanding the physical safety, security, and privacy concerns of people with visual impairments. *IEEE Internet Computing*, 21(3):56–63, 2017.

[2] T. Ahmed, P. Shaffer, D. Connelly, K. Crandall, and A. Kapadia. Addressing physical safety, security, and privacy for people with visual impairments. In *Proceedings of the Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*, 2016.

[3] T. Akter, T. Ahmed, A. Kapadia, and M. Swaminathan. Shared privacy concerns of the visually impaired and sighted bystanders with camera-based assistive technologies. *ACM Transactions on Accessible Computing*, 15(2):1–33, 2022.

[4] T. Akter, B. Dosono, T. Ahmed, A. Kapadia, and B. Semaan. "I am uncomfortable sharing what I can't see": Privacy concerns of the visually impaired with camera based assistive applications. In *Proceedings of the 29th USENIX Security Symposium*, 2020.

[5] J. A. Bargh, R. N. Bond, W. J. Lombardi, and M. E. Tota. The additive nature of chronic and temporary sources of construct accessibility. *Journal of Personality and Social Psychology*, 50(5):869–878, 1986.

[6] C. Barwulor, A. McDonald, E. Hargittai, and E. M. Redmiles. "Disadvantaged in the American-dominated internet": Sex, work, and technology. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021.

[7] R. Bellini. Paying the price: When intimate partners use technology for financial harm. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023.

[8] R. Bellini, E. Tseng, N. Warford, A. Daffalla, T. Matthews, S. Consolvo, J. P. Woelfer, P. G. Kelley, M. L. Mazurek, D. Cuomo, N. Dell, and T. Ristenpart. SoK: Safer Digital-Safety Research Involving At-Risk Users. In *Proceedings of the IEEE Symposium on Security and Privacy*, 2024.

[9] R. Benjamin. *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity, 2019.

[10] R. Bhalerao, V. Hamilton, A. McDonald, E. M. Redmiles, and A. Strohmayer. Ethical practices for security research with at-risk populations. In *2022 IEEE European Symposium on Security and Privacy Workshops*, 2022.

[11] M. Bishop, L. Drevin, L. Futcher, W. Leung, N. Miloslavskaya, E. Moore, J. Ophoff, and S. von Solms. A brief history and overview of WISE. In L. Drevin, N. Miloslavskaya, W. Leung, and S. von Solms, editors, *Information Security Education for Cyber Resilience*, pages 3–9. Springer, 2021.

[12] M. Blanz. Accessibility and fit as determinants of the salience of social categorizations. *European Journal of Social Psychology*, 29:43–74, 1999.

[13] J. Buolamwini and T. Gebru. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency (Proceedings of Machine Learning Research)*, 2018.

[14] P. D. Chowdhury and K. V. Renaud. 'Ought' should not assume 'Can' ... Basic Capabilities in Cybersecurity to Ground Sen's Capability Approach. In *Proceedings of the 2023 New Security Paradigms Workshop*, 2023.

[15] J. Cleland-Huang. How well do you know your personae non gratae? *IEEE Software*, 31(4):28–31, 2014.

[16] S. Costanza-Chock. *Design Justice: Community-Led Practices to Build the Worlds we Need*. The MIT Press, 2020.

[17] K. W. Crenshaw. Demarginalizing the intersection of race and sex: A Black feminist critique of antidiscrimination doctrine. *University of Chicago Legal Forum*, pages 139–168, 1989.

[18] C. Criado Perez. *Invisible Women: Exposing Data Bias in a World Designed for Men*. Abrams Books, 2019.

[19] A. Czeskis, I. Dermendjieva, H. Yapit, A. Borning, B. Friedman, B. Gill, and T. Kohno. Parenting from the pocket: Value tensions and technical directions for secure and private parent-teen mobile safety. In *Symposium On Usable Privacy and Security (SOUPS)*, 2010.

[20] M. del Bosque. Facial Recognition Bias Frustrates Black Asylum Applicants to US, Advocates Say. *The Guardian*, 2023. https://www.theguardian.com/us-news/2023/feb/08/us-immigration-cbp-one-app-facial-recognition-bias.

[21] T. A. Denning, B. Friedman, and T. Kohno. *Security Cards: A Security Threat Brainstorming Toolkit*. University of Washington, 2013.

[22] J. C. Deska, E. P. Lloyd, and K. Hugenberg. Facing humanness: Facial width-to-height ratio predicts ascriptions of humanity. *Journal of Personality and Social Psychology*, 114(1):75–94, 2018.

[23] P. G. Devine, P. S. Forscher, A. J. Austin, and W. T. Cox. Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology*, 48(6):1267–1278, 2012.

[24] P. G. Devine, P. S. Forscher, W. T. Cox, A. Kaatz, J. Sheridan, and M. Carnes. A gender bias habit-breaking intervention led to increased hiring of female faculty in STEMM departments. *Journal of Experimental Social Psychology*, 73:211–215, 2017.

[25] T. Devos and M. R. Banaji. American = white? *Journal of Personality and Social Psychology*, 88(3):447–466, 2005.

[26] B. Friedman and D. G. Hendry. *Value Sensitive Design: Shaping Technology with Moral Imagination*. The MIT Press, 2019.

[27] B. Friedman, P. H. Kahn Jr., and A. Borning. Value sensitive design: Theory and methods. Technical report, University of Washington, 2002.

[28] A. Frik, L. Nurgalieva, J. Bernd, J. S. Lee, F. Schaub, and S. Egelman. Privacy and security threat models and mitigation strategies of older adults. In *USENIX Symposium on Usable Privacy and Security (SOUPS)*, 2019.

[29] C. Geeng, M. Harris, E. M. Redmiles, and F. Roesner. "Like lesbians walking the perimeter": Experiences of u.s. lgbtq+ folks with online security, safety, and privacy advice. In *Proceedings of the 31st USENIX Security Symposium*, 2022.

[30] A. K. Ghosh, K. Badillo-Urquiola, S. Guha, J. J. LaViola Jr, and P. J. Wisniewski. Safety vs. surveillance: What children have to say about mobile apps for parental control. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018.

[31] E. Goffman. *Stigma: Notes on the Management of Spoiled Identity*. Prentice-Hall, 1963.

[32] T. Guberek, A. McDonald, S. Simioni, A. H. Mhaidli, K. Toyama, and F. Schaub. Keeping a low profile? technology, risk and privacy among undocumented immigrants. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018.

[33] E. V. Hall, A. V. Hall, A. D. Galinsky, and K. W. Phillips. MOSAIC: A model of stereotyping through associated and intersectional categories. *Academy of Management Review*, 44(3):643–672, 2019.

[34] V. Hamilton, H. Barakat, and E. M. Redmiles. Risk, resilience and reward: Impacts of shifting to digital sex work. In *Proceedings of the ACM on Human-Computer Interaction*, 2022.

[35] E. T. Higgins. Knowledge activation: Accessibility, applicability, and salience. In E. T. Higgins and A. W. Kruglanski, editors, *Social Psychology: Handbook of Basic Principles*, pages 133–168. Guilford Press, 1996.

[36] D. Hornung, C. Müller, I. Shklovski, T. Jakobi, and V. Wulf. Navigating relationships and boundaries: Concerns around ict-uptake for elderly people. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017.

[37] R. Jeong and S. Chiasson. 'Lime', 'Open Lock', and 'Blocked': Children's perception of colors, symbols, and words in cybersecurity warnings. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020.

[38] L. Kohnfelder and P. Garg. *The Threats to Our Products*. Microsoft Interface, 1999.

[39] T. Kohno. *Background and Context for the Our Reality Novella*. 2021.

[40] T. Kohno. *Our Reality: A Novella*. 2021.

[41] T. Kohno and B. D. Johnson. Science fiction prototyping and security education: Cultivating contextual and societal thinking in computer security education and beyond. In *Proceedings of the 42nd ACM Technical Symposium on Computer Science Education*, 2011.

[42] P. C. Kumar, M. Chetty, T. L. Clegg, and J. Vitak. Privacy and security considerations for digital technology use in elementary schools. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019.
[43] E. Lastdrager, I. C. Gallardo, P. Hartel, and M. Junger. How effective is anti-phishing training for children? In *Proceedings of the Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*, 2017.
[44] A. Lerner, H. Y. He, A. Kawakami, S. C. Zeamer, and R. Hoyle. Privacy and activism in the transgender community. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020.
[45] G. Liveley. *Stories of Cyber Security Combined Report.* 2022.
[46] E. P. Lloyd, K. Hugenberg, A. R. McConnell, J. W. Kunstman, and J. C. Deska. Black and white lies: Race-based biases in deception judgments. *Psychological Science*, 28(8):1125–1136, 2017.
[47] K. B. Maddox and S. Gray Chase. Manipulating subcategory salience: Exploring the link between skin tone and social perception of Blacks. *European Journal of Social Psychology*, 34:533–546, 2004.
[48] J. K. Maner, S. L. Miller, J. H. Moss, J. L. Leo, and E. A. Plant. Motivated social categorization: Fundamental motives enhance people's sensitivity to basic social categories. *Journal of Personality and Social Psychology*, 103(1):70–83, 2012.
[49] A. McDonald, C. Barwulor, M. L. Mazurek, F. Schaub, and E. M. Redmiles. "It's stressful having all these phones": Investigating sex workers' safety goals, risks, and practices online. In *Proceedings of the 30th USENIX Security Symposium*, 2021.
[50] B. McNally, P. Kumar, C. Hordatt, M. L. Mauriello, S. Naik, L. Norooz, A. Shorter, E. Golub, and A. Druin. Co-designing mobile online safety applications with children. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018.
[51] A. R. McNeill, L. Coventry, J. Pywell, and P. Briggs. Privacy considerations when designing social network systems to support successful ageing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017.
[52] E. McReynolds, S. Hubbard, T. Lau, A. Saraf, M. Cakmak, and F. Roesner. Toys that listen: A study of parents, children, and internet-connected toys. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017.
[53] N. R. Mead, F. Shull, K. Vemuru, and O. Villadsen. *A Hybrid Threat Modeling Method.* Carnegie Mellon University, 2018.
[54] J. Mirkovic, M. Dark, W. Du, G. Vigna, and T. Denning. Evaluating cybersecurity education interventions: Three case studies. *IEEE Security & Privacy*, 13(3):63–69, 2015.
[55] C. Moser, T. Chen, and S. Y. Schoenebeck. Parents' and children's preferences about parents sharing about children on social media. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017.
[56] J. Nicholson, L. Coventry, and P. Briggs. "If it's important it will be a headline": Cybersecurity information seeking in older adults. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019.
[57] K. Owens, A. Alem, F. Roesner, and T. Kohno. Electronic monitoring smartphone apps: An analysis of risks from technical, human-centered, and legal perspectives. In *31st USENIX Security Symposium*, 2022.
[58] K. Owens, C. Cobb, and L. Cranor. "You gotta watch what you say": Surveillance of communication with incarcerated people. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021.
[59] E. L. Paluck and D. P. Green. Prejudice reduction: What works? a review and assessment of research and practice. *Annual Review of Psychology*, 60:339–367, 2009.
[60] Paul G. Allen School of Computer Science and Engineering. Allen school demographics. 2022. https://www.cs.washington.edu/diversity/demographics.
[61] S. Perkowitz. The bias in the machine: Facial recognition technology and racial disparities. *MIT Schwarzman College of Computing*, 2021. https://mit-serc.pubpub.org/pub/bias-in-machine/release/1.
[62] J. Petelka, M. Finn, F. Roesner, and K. Shilton. Principles Matter: Integrating an Ethics Intervention into a Computer Security Course. In *53rd ACM Technical Symposium on Computer Science Education (SIGCSE)*, 2022.
[63] V. Purdie-Vaughns and R. P. Eibach. Intersectional invisibility: The distinctive advantages and disadvantages of multiple subordinate-group identities. *Sex Roles*, 59:377–391, 2008.
[64] F. Roesner and T. Kohno. Security and privacy for augmented reality: Our 10-year retrospective. In *VR4Sec: 1st International Workshop on Security for XR and XR for Security*, 2021.
[65] S. Sannon and A. Forte. Privacy research with marginalized groups: What we know, what's needed, and what's next. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2), Nov. 2022.
[66] A. K. Sesko and M. Biernat. Prototypes of race and gender: The invisibility of Black women. *Journal of Experimental Social Psychology*, 46(2):356–360, 2010.
[67] F. Sharevski and A. Zeidieh. "I Just Didn't Notice It:" Experiences with Misinformation Warnings on Social Media amongst Users Who Are Low Vision or Blind. In *Proceedings of the 2023 New Security Paradigms Workshop*, 2023.
[68] N. Shawl and C. Ward. *Writing the Other: A Practical Approach.* Aqueduct Press, 2005.
[69] L. Simko, A. Lerner, S. Ibtasam, F. Roesner, and T. Kohno. Computer security and privacy for refugees in the united states. In *2018 IEEE Symposium on Security and Privacy*, 2018.
[70] J. Slupska, S. D. Dawson Duckworth, L. Ma, and G. Neff. Participatory threat modeling: Exploring paths to reconfigure cybersecurity. In *Extended abstracts of the 2021 CHI conference on human factors in computing systems*, 2021.
[71] K. Thomas, D. Akhawe, M. Bailey, D. Boneh, E. Bursztein, S. Consolvo, N. Dell, Z. Durumeric, P. G. Kelley, D. Kumar, D. McCoy, S. Meiklejohn, T. Ristenpart, and G. Stringhini. SoK: Hate, Harassment, and the Changing Landscape of Online Abuse. In *Proceedings of the IEEE Symposium on Security and Privacy*, 2021.
[72] E. Tseng, M. Sabet, R. Bellini, H. K. Sodhi, T. Ristenpart, and N. Dell. Care infrastructures for digital security in intimate partner violence. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 2022.
[73] N. Warford, T. Matthews, K. Yang, O. Akgul, S. Consolvo, P. G. Kelley, N. Malkin, M. L. Mazurek, M. Sleeper, and K. Thomas. SoK: A Framework for Unifying At-Risk User Research. In *Proceedings of the IEEE Symposium on Security and Privacy*, 2022.
[74] C. Weir, I. Becker, J. Noble, L. Blair, M. A. Sasse, and A. Rashid. Interventions for long-term software security: Creating a lightweight program of assurance techniques for developers. *Software: Practice and Experience*, 50(3):275–298, 2020.
[75] T. Yip, C. S. L. Cheah, L. Kiang, and G. C. Nagayama Hall. Rendered invisible: Are Asian Americans a model or a marginalized minority? *American Psychological Association*, 76(4):575–581, 2021.
[76] J. Zhao, G. Wang, C. Dally, P. Slovak, J. Edbrooke-Childs, M. Van Kleek, and N. Shadbolt. 'I make up a silly name': Understanding children's perception of privacy risks online. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019.
[77] Y. Zou, A. McDonald, J. Narakornpichit, N. Dell, T. Ristenpart, K. A. Roundy, F. Schaub, and A. Tamersoy. The role of computer security customer support in helping survivors of intimate partner violence. In *Proceedings of the 30th USENIX Security Symposium*, 2021.
[78] M. E. Zurko. User-centered security: Stepping up to the grand challenge. In *Proceedings of the 21st Annual Computer Security Applications Conference*, 2005.
[79] M. E. Zurko and R. T. Simon. User-centered security. In *New Security Paradigms Workshop*, 1996.

A THREAT MODELING EXERCISE

A.1 Time 1

Your threat modeling target is the following technology: An augmented reality (AR) headset. Many companies have begun to develop augmented/mixed/virtual reality headsets, such as the Oculus (from Meta) or the HoloLens (from Microsoft, depicted below). Augmented reality (AR) technologies allow people to interact with virtual content overlaid on their perception of the physical world – for example, AR applications might label objects or people in the world, support immersive games like Pokemon Go, show directions overlaid on the physical world, and much more. But, as with any technology, there are potential security and privacy concerns.

Before discussing with anyone, please fill out and go ahead and submit answers to the following questions:

(1) What do you think are the **security goals** of the AR headset described in class and shown above? What **assets** must be protected?
(2) Who are the **adversaries** who might try to attack this AR headset? What might be the **attacker's goals**? What potential **threats or vulnerabilities** do you see?

A.2 Time 2

A.2.1 Control Prompt. Research shows that sometimes when people consider a question a second time, they come up with different responses or think differently about a problem. Take a few minutes to consider these questions a second time, and submit your answers again.

(1) What do you think are the **security goals** of the AR headset described in class and shown above? What **assets** must be protected?

(2) Who are the **adversaries** who might try to attack this AR headset? What might be the **attacker's goals**? What potential **threats or vulnerabilities** do you see?

A.2.2 Salience Intervention Prompt. Sometimes engineers default to unintentionally designing for some populations and not others. For example:

- A classic example is the dummies used in automobile crash tests: they were designed to match the anatomy of a 70kg adult man, thereby excluding from consideration much of the population.
- Another example is face recognition technologies, which have been in the news because of a failure to design for racial and gender diversity.
- Yet another example is the assumption that smartphones have only a single user, which may not hold for parents sharing devices with their children, or people in lower-income or non-US contexts.

To avoid accidentally only considering some "default" stakeholder groups, try to be creative and think about populations that engineers might not normally think about. Consider and answer again the following questions.

(1) What do you think are the **security goals** of the AR headset described in class and shown above? What **assets** must be protected?
(2) Who are the **adversaries** who might try to attack this AR headset? What might be the **attacker's goals**? What potential **threats or vulnerabilities** do you see?

B OPT-OUT EMAIL

Instructors sent their class an email to allow them to opt-out of having their data included in the studies. The emails were sent out at different times across the two quarters, leading to slightly different wording (e.g., future tense vs present tense). For instance, in the future tense version of this email sent before the finalization of grades, students were explicitly told that opting out would not impact their grades. We include the present tense version of the email below, which was sent after the quarter was over and grades were finalized.

Hi everyone,

As part of our research related to computer security education, we tried out a few different versions of in-section threat modeling activities (one in Section 1 and one in Section 9). Our research goals are focused on developing interventions to help people doing threat modeling or security analyses consider a diverse range of possible stakeholders beyond a potential "default persona".

Now that the quarter is finished, we plan to study your in-section activity responses as part of our research. Some important things to know:

- If you wish to opt out and not allow us to use your in-section activity responses as part of our research, you are able to do so. Specifically, let me know by replying to me if you'd like to opt out of having your (anonymized) week 1 and week 9 in-section activities included in a research study.
- For those who don't opt out, there is an option to provide some demographic information. Optionally, fill out this demographic survey [link].
- We will remove identifying information from your in-section activities and your survey responses. The rest of the research team (other than me) will only see new, numeric identifiers for each student. We will discard the mapping from student to identifier after applying it (so even I will no longer see identifiers when we analyze the data).
- This study was approved by the UW's Human Subjects Research Review Board (aka IRB) [link to IRB website].
- While different sections saw slightly different material in the in-section activities, all sections were given all versions of the educational material by the end of the quarter.

If you have any questions, now or later, please don't hesitate to let me know. Thanks!

C STAKEHOLDER CATEGORIES

Table 2 summarizes and categories the stakeholders produced by our expert panel.

Stakeholder Category	Examples
Activists/politically involved citizens	People going to protests Activists Voters
AR 3rd party entities	3rd party apps 3rd party developer Application designer Co-located apps Internet provider, edge computing Mobile network operators Broadband companies Investors
AR developers/designers	System developers/designers Shareholders of the companies Platform designer AR infrastructure operators
AR Regulators	EU pro-regulatory entity Civil society NGOs - anti/pro-tech Standard-setting organizations (ISO)
Bystanders	Bystanders Pedestrians People with relationships to users
Celebrity and social accounts	Celebrities Content creators Digital streamers Influencers Social media account managers Fans
Default	End user (abstract) Cultural default
Employees/patients	Entrepreneurs Employees/company office employees Medical patients
Employers and companies	Employer of a company using AR/VR Employers Advertisers Schools Insurance companies Management class
Entities monitoring others	Parents Teachers Hackers School bully Former/current partner Law enforcement Police
Government entities	Governments International governments/ entities (UN) Policymakers Politicians Military
Non-U.S. citizens	Undocumented immigrants Non-native English speakers Non-citizen residents of a country
People without access to the technology	People with low levels of tech expertise People who can't afford AR headset Non-users (by choice, or because inaccessible)
People with stigmatized social identities	Religious person (e.g., with religious attire) People with physical disabilities People with cognitive disabilities Blind people/people with impaired vision Gender minorities Racial minorities
Specific app users	Gamers Shoppers Drivers navigating
Targets of hate, harassment, and abuse	Domestic abuse victims/ survivors Victim of bully Targets of hate and harassment
Vulnerable age groups/ people unable to consent	Children Teenagers Students Older adults People suspected of crimes, under electronic monitoring
Vulnerable workers	Workers in a factory Gig workers Low wage workers Child care workers Journalists Sex workers

Table 2: Stakeholders generated by the expert panel (right column), clustered into 18 distinct categories (left column).