Deceptive Design Patterns in Safety Technologies: A Case Study of the Citizen App

Ishita Chordia
The Information School, University of Washington
Seattle, USA
ichordia@uw.edu

Lena-Phuong Tran
Human-Centered Design and Engineering, University of Washington
Seattle, USA
lpktran@uw.edu

Tala Tayebi
The Information School, University of Washington
Seattle, USA
tayebit@uw.edu

Emily Parrish
The Information School, University of Washington
Seattle, USA
eparrish@uw.edu

Sheena Erete
College of Information Studies, The University of Maryland
College Park, USA
serete@umd.edu

Jason Yip
The Information School, University of Washington
Seattle, USA
jcyip@uw.edu

Alexis Hiniker
The Information School, University of Washington
Seattle, USA
alexisr@uw.edu

ABSTRACT

Deceptive design patterns (known as dark patterns) are interface characteristics which modify users' choice architecture to gain users' attention, data, and money. Deceptive design patterns have yet to be documented in safety technologies despite evidence that designers of safety technologies make decisions that can powerfully influence user behavior. To address this gap, we conduct a case study of the Citizen app, a commercially available technology which notifies users about local safety incidents. We bound our study to Atlanta and triangulate interview data with an analysis of the user interface. Our results indicate that Citizen heightens users' anxiety about safety while encouraging the use of profit-generating features which offer security. These findings contribute to an emerging conversation about how deceptive design patterns interact with sociocultural factors to produce *deceptive infrastructure*. We propose the need to expand an existing taxonomy of harm to include *emotional load* and *social injustice* and offer recommendations for designers interested in dismantling the deceptive infrastructure of safety technologies.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**.

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04.
https://doi.org/10.1145/3544548.3581258

KEYWORDS

dark patterns, safety, dark infrastructure, manipulative design, deceptive design, crime, community safety, fear, anxiety, safety technologies

ACM Reference Format:
Ishita Chordia, Lena-Phuong Tran, Tala Tayebi, Emily Parrish, Sheena Erete, Jason Yip, and Alexis Hiniker. 2023. Deceptive Design Patterns in Safety Technologies: A Case Study of the Citizen App. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3544548.3581258

1 INTRODUCTION

In 2010, Brigunull coined the term "dark patterns" to describe how interface design can be used to undermine user agency and manipulate user decision-making [18]. He describes dark patterns as "*tricks used in websites and apps that make you do things that you didn't mean to, like buying or signing up for something*" [18]. Examples of dark patterns originally identified by Brignull include "Sneak Into Basket," when a site sneaks additional items into the shopping cart without your consent, and "Disguised Ads," where advertisements are disguised as content or navigation. Since 2010, research on dark patterns has grown substantially and has evolved to include both explicitly manipulative "tricks" and lighter nudges which, at scale, can cause harm to both users and society [75]. Recent literature has identified dark patterns in e-commerce [74], digital privacy [20, 32], social media [73, 79], ubiquitous computing [50], robotics [64], and gaming [111] domains. The terminology has also evolved [18, 87]- in the remainder of the paper, we refer to dark patterns as "deceptive design patterns" to avoid equating the racialized term "dark" with problematic behavior.

Understanding deceptive design patterns and how they operate in different contexts is increasingly important, as deceptive design patterns are now pervasive and, for example, employed in the vast majority (95%) of apps on the Google Play Store [37]. Researchers, however, have yet to document the existence of deceptive design patterns in safety technologies. "Safety technology" refers to any digital technologies used for the purpose of increasing user safety. Designers of commercial safety technologies make decisions that powerfully influence users' behavior [44]. Prior literature, for example, has documented how safety technologies can influence users' levels of civic engagement [41, 44], their interactions with other members of their communities [44, 69, 70], the social norms of the neighborhood [63, 90], and individuals' feelings of safety [14, 58]. Safety technologies can also impact individuals who are not users of those technologies by contributing to racial profiling [44, 70] and online racism [109].

Given that designers of safety technologies make decisions that can have consequences for both users and non-users of these technologies and can shape both online and offline behavior, we were curious about how these decisions may be influenced by profit motives. We conduct a case study [76] of the Citizen app, a commercially available location-based crime alert technology that notifies users about local incidents related to public safety. We interview fifteen users of the Citizen app who live in Atlanta, a racially diverse mid-sized city in the Southern portion of the United States. To understand how deceptive design patterns influence the user experience, we triangulate the interview data with an interface analysis of the app.

We ask:

- **RQ1:** How, if at all, does the design of the Citizen interface reflect known deceptive design patterns?
- **RQ2:** How do these designs affect the user experience?

We find that Citizen employs a collection of user interface elements that together raise the salience of safety incidents, emphasizing the extent to which reported incidents pose a threat to the user. The app further presents itself as a solution to danger, leveraging a collection of common deceptive design patterns to exert purchase pressure on the user and encourage data disclosure. Participants' experiences aligned with this feature analysis. They voiced an appreciation for receiving hyper-local, real-time safety information that helped them navigate risk, but many also reported that the app's information-sharing practices increased fear and encouraged dependence on the app. Furthermore, users explained that Citizen influenced their offline behavior, including the neighborhoods they visited and their interactions with Black and unhoused individuals perceived to be dangerous.

Deceptive infrastructure (sometimes known as dark infrastructure) refers to the interactions between deceptive design patterns and larger social, psychological, and cultural factors that together undermine user agency at scale [108]. Our study contributes to an emerging conversation on deceptive infrastructure by demonstrating how deceptive design patterns, human biases, and sociocultural contexts interact to produce harm for both users and non-users of the Citizen app. Deceptive design patterns interact with attentional bias to create anxiety for users and interact with negative cultural stereotypes to disproportionately harm vulnerable and historically marginalized populations. In light of these results, we identify *emotional load* and *social injustice* as two forms of harm perpetuated by deceptive design patterns that are yet to be documented [75]. We additionally offer four concrete suggestions to designers of safety technologies who are interested in dismantling the deceptive infrastructure produced by existing safety technologies.

2 RELATED WORK

2.1 Deceptive Design Patterns in HCI

A recent review of the literature on deceptive design patterns in HCI finds that while there are many different definitions, a uniting characteristic is that deceptive design patterns all *"modify the underlying choice architecture for users"* [75, p.9]. Deceptive design patterns grew out of manipulative practices in retail, research on persuasive design, and digital marketing as a way for companies to gain users' attention, data, and money [80]. Deceptive design patterns use language, emotion, color, style, and cognitive biases to undermine user agency [74, 75]. They are pervasive on online platforms and have been documented in the vast majority (95%) of apps on the Google Play Store [37], including e-commerce [74], gaming [111], and social media platforms [73, 79]. Examples of common deceptive design patterns include "Infinite Scrolling" [79] where new content automatically loads as users scroll the page and "Hard to Cancel" subscriptions [74].

Deceptive design patterns can be highly effective in manipulating user behavior [72, 83]. Prior research has found that American consumers are twice as likely to sign up for a premium theft protection service when presented with a mild deceptive design pattern and four times as likely to sign up when presented with an aggressive deceptive design pattern compared to users who are shown a neutral interface [83]. Calo and Rosenblat argue that digital technologies are uniquely effective at influencing user behavior because of their ability to capture and store information about users, their ability to architect virtually every aspect of the platforms, and their ability to translate insight about user behavior into design [21, 22]. Furthermore, the emergence of online markets, such as digital sharing economies, presents new opportunities for companies to manipulate users by modifying the choice architecture of both sellers (e.g. Uber drivers) and buyers (e.g. riders) [21, 22].

Deceptive design patterns can diminish user wellbeing through financial loss, invasion of privacy, and cognitive burdens [75]. Schull and others have found that social media platforms employ addictive deceptive design patterns, such as infinite scroll or Youtube's autoplay, that rely on a variable reward that mimics strategies used by the gambling industry [67, 97], and prior work has even documented the prevalence of deceptive design patterns in mobile applications for children [87]. The impact of deceptive design patterns, however, is not limited to individual users. Mathur and colleagues discuss the potential for deceptive design patterns to also impact collective welfare, by decreasing trust in the marketplace and by contributing to unanticipated societal consequences [75]. They point to Cambridge Analytica's use of personal data, collected with the help of deceptive design patterns on Facebook, to influence the 2016 U.S. presidential election as an example.

Given that deceptive design patterns are effective at manipulating user behavior and can negatively impact both individual and

collective welfare, it is important to understand how they operate in different contexts. In the present study, we investigate the incidence and influence of deceptive design patterns in safety technologies, filling a gap in the field.

2.2 Safety Technologies in HCI

There is a rich body of work on safety technologies in HCI that spans more than a decade. Much of the early literature sought to design technologies to reduce individuals' risk of victimization [14, 58, 99, 103]. This work was influenced by victimization theory from criminology which views victims and offenders as rational actors who use the information they have to assess their risk of being victimized or caught, respectively [68]. Digital technologies inspired by this perspective sought to provide users with information that would lower their chance of victimization. For example, Blom and colleagues designed a mobile application that allowed women to view and label spaces as "safe" or "unsafe" [14], and Shah prototyped CrowdSafe, which shared location-based crime information and traffic navigation guidance with users [99]. These technologies focused on decreasing individuals' risk.

In contrast to the victimization theory, the social control theory focused on the community and the informal and formal controls in place to deter crime [68, 92]. Digital technologies drawing from this theory emphasized the importance of not only sharing information with individuals, but also supporting community engagement, collaboration, and problem-solving [58, 69]. Researchers in HCI studied neighborhood listservs [41, 44, 69] and social media [54, 56, 90, 91, 109, 112] to understand how to increase collaboration between citizens and local authorities [91, 112], encourage civic engagement [41, 44], support user engagement and information sharing [19, 58, 69], and decrease individuals' fear of crime [14, 15, 58].

The most recent work examining safety technologies in HCI has investigated the potential for safety technologies to perpetuate harm against historically marginalized populations. For example, empirical work studying online communication on local neighborhood listservs and Nextdoor find that these platforms serve as spaces for online negotiations of "suspicious behavior" that can lead to increased policing and surveillance of people of color [63, 70, 71, 78]. On Reddit, ambiguous and passive policies towards racist comments that are focused more on protecting Reddit's image and user engagement lead to both new and old racism in discussions of safety [109]. Researchers have documented similar patterns of racism, policing, and surveillance on other apps where users organize around and discuss community safety, such as WhatsApp [78] and Amazon Neighbors [17]. A study analyzing product reviews and promotional material of Citizen, Nextdoor, and bSafe [59] found that companies encourage users to surveil members of their communities, leading users to express fear and racist beliefs. Collectively, this research suggests a need to investigate the role that design plays in perpetuating harm against historically marginalized populations. Sociologist Rahim Kurwa explains that such work is critical because surveillance and policing "relies[rely] on de-racialized governing narratives of safety that nevertheless have racist implementation and results" [63, pg.114].

Designers of safety technologies make decisions that shape users' individual and collective behavior. Furthermore, these technologies can have harmful and far-reaching consequences. By studying deceptive design patterns, we can begin to understand the factors that motivate these influential design decisions.

3 CASE STUDY DESIGN

We employed a case study method [76] to understand how deceptive design patterns influence the user experience of safety applications. We investigated a single case, the Citizen app, and bound our study to Atlanta users and their experience with the app from 2021 to 2022. This work is a case study because of our in-depth, holistic description and analysis of a bounded phenomenon [76].

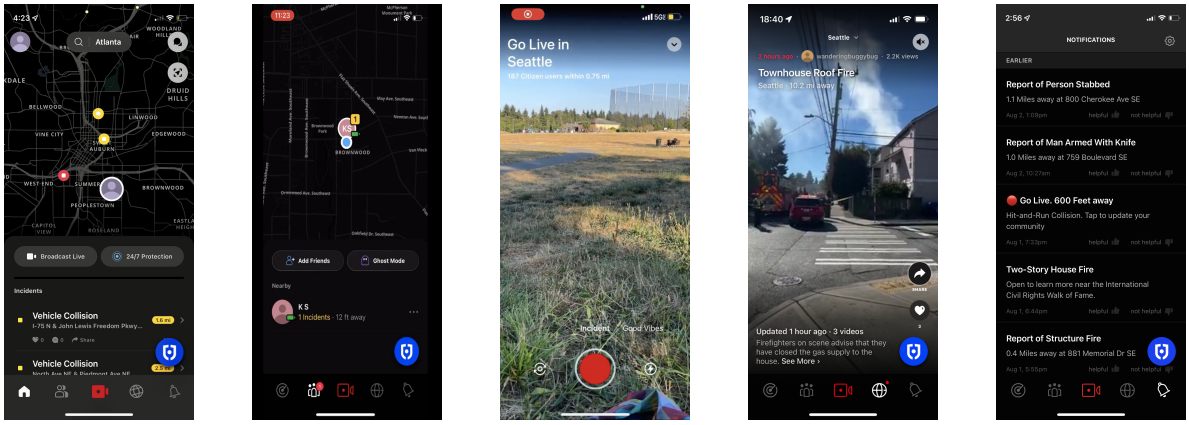
For the single case to have power, the selection of the case needs to be strategic [47]. We selected Citizen because we see it as an *extreme* case [110]. Citizen deviates from other safety technologies in its profit model because it does not sell advertisements nor does it sell user data [26]. Rather, Citizen's premium feature connects users to Citizen employees who monitor a user's surroundings; this is the only way Citizen currently generates revenue. We chose Citizen for our case because we hypothesized that this business model may have unique implications on the design of the application. At the same time, because Citizen has many of the same features as other safety technologies, including the ability to view and discuss safety incidents, receive alerts about safety incidents, and view location-specific data, we hypothesized that our findings may reveal insights about other safety technologies as well.

We triangulated data from two sources [110]. We first conducted user interviews and asked participants about the influence of individual features to allow evidence of deceptive design patterns to emerge organically. We then conducted a researcher-led review of the user interface to identify known deceptive design patterns. In the following sections, we give context for our case and describe our process for collecting and analyzing data.

3.1 Contextual Background

3.1.1 Atlanta Context. We chose to geographically bound our investigation to users of Citizen that live in and around Atlanta. We chose a city in which Citizen was available, where crime was a concern, and where the authors had access to online neighborhood groups and/or local Facebook pages for recruitment. Prior research suggests that safety technologies are used differently by different communities [41, 44], and we hoped that by geographically bounding our investigation, we may better see patterns in individual behavior. It is important to note, however, that the experience of users in Atlanta is not necessarily representative of users from other USA cities.

Atlanta is a racially diverse city in the Southeastern portion of the United States. According to the 2021 Census [6], Black people make up the largest percentage of the city (51%), followed by White people (40.9%), and Asians (4.4%). Once considered a "Black Mecca," Atlanta's racial demographics have, however, changed drastically in the last decade. For the first time since the 1920s, the Black population has been declining while the White population has been growing [36]. This racial shift can be attributed to the recent onset of gentrification, as well as the population growth of the city [65].



(a) (b) (c) (d) (e)

Figure 1: Citizen is made up of five main tabs: a) Home Page tab, b) Safety Network tab, c) Live Broadcast tab, d) Newsfeed tab, e) Notifications tab.

Additionally, in 2019, Atlanta had the second largest inequality gap in the country [9], with 20% of the population living below the poverty line [10]. A survey collected by the City Continuum of Care counted roughly 3,200 unhoused individuals in 2020, 88% of whom were Black [48].

In the early 2000s, Atlanta had one of the highest rates of violent crime in the country. Although crime rates have largely decreased in the 2010s, violent crimes such as homicides, aggravated assaults, and shooting incidents have gone up since 2017 [85]. Between 2019 and 2021, homicide increased 54% and aggravated assaults by 23% [31, 53]. This is consistent with numbers in large cities across the country who have all experienced a surge in violent crime during the COVID-19 pandemic [7]. In addition to an increase in violent crime, Atlanta has faced outrage and protests due to a number of high-visibility murders of Black people at the hands of police and White vigilantes [46, 84].

3.1.2 Citizen Context. Citizen is a location-based crime alert platform that notifies users about local incidents which can affect public safety [55]. Citizen was originally released in 2016 as Vigilante, a platform where users could develop vigilante-style networks to protect themselves from potential offenders. After being banned from the Apple App Store for its potential to incite violence, parent company Sp0n re-branded and re-released the platform as Citizen in 2017 [55]. The mission of the app, as reported on its website in August 2022, reads: "We live in a world where people can access information quickly, share effortlessly, and connect easily — but we have yet to see the power of bringing people together to watch out for each other. At Citizen, we're developing cutting-edge technology so you can take care of the people and places you love" [27].

Citizen's custom-built AI algorithm listens to first-responder radio transmissions. From these raw feeds, the AI algorithm automatically processes radio clips and extracts keywords. A Citizen analyst listening to the 911 dispatch then writes a short incident notification, which may be sent to users as an alert [13]. These incidents are supplemented with crowdsourced user videos, which are reviewed by the company's moderators before appearing on the app. The Citizen FAQ reports that they include "major incidents that are in progress, or ones that we assess could affect public safety" [24]. The radius around which a user will receive notifications varies based on a number of factors, including the "nature of the incident and the population density of the area" [23].

The basic version of the app is free, does not have ads, and CEO Frame says it does not sell or share user data [13]. However, Citizen is currently facing pressure from venture capitalists backing the platform to monetize and is experimenting with premium features, such as "Citizen Protect," which allows users to contact company employees to virtually monitor their surroundings and dispatch emergency responders [8].

There are five tabs that users can interact with in the app and Figure 1 shows screenshots of each tab. The five tabs are: 1) the Home tab, which displays a map with the user's current location and nearby incidents as well as a list of nearby incidents; 2) the Safety Network tab, which displays a map of the user's friends' current locations. This tab also display the safety incidents near each friend, the distance from each friend to the user, and the battery life remaining on each friend's mobile device; 3) the Broadcast tab, which allows users to record live videos. They can choose between two types of live videos: "Incidents" or "Good Vibes" with the app defaulting to "Incidents"; 4) the Newsfeed tab, which shows live videos captured by users. Tapping into a video takes users to a page with more information about the incident, including additional video clips (if available), a list of updates, comments and reactions from other users, and the address on a map. In addition to local incidents, users can also choose to view incidents in other major cities or a "global" category; 5) the Notifications tab lists a history of all reported incidents since the user joined the app.

As of January 2022, Citizen is released in 60 cities or metro areas [25]. Citizen was made available in Atlanta in October 2020, and as of November 2020, was reported to have over 17,000 users [16].

Table 1: Demographic Characteristics of Study Participants

Participant ID	Race	Age	Gender	Length of Time Using Citizen	Time Spent on Citizen Each Week
P1	White	25-34	Female	About 4 months	30 minutes-1 hour, maybe more
P2	White	45-54	Female	7 months	10 minutes
P3	Black	25-34	Female	5 weeks	60 minutes
P4	White	25-34	Female	2 months	About an hour
P5	Undisclosed	35-44	Undisclosed	Over a year	5 minutes
P6	White and Native American	35-44	Female	3-6 months	30 minutes or so
P7	Asian or Pacific Islander	35-44	Male	6 months	30 minutes
P8	Hispanic or Latino/a	35-44	Male	6 months	30 minutes
P9	White	25-34	Female	1.5 years	10-20 minutes
P10	White	65-74	Male	3 months	Only when notified
P11	White	35-44	Male	1.5 years	1.5 hours
P12	White	35-44	Female	2 years	20 minutes
P13	Black	18-24	Male	Undisclosed	12 hours
P14	Black	25-34	Male	6 months	Undisclosed
P15	Black	18-24	Male	2 years	10-15 minutes

3.2 Data Collection

3.2.1 User Interviews. Two members of the research team conducted fifteen semi-structured Zoom [113] interviews with Citizen users who live in and around Atlanta. The first twelve interviews were conducted between September and October 2021, and an additional three interviews were conducted in June and July 2022 targeting people of color so that our findings would better reflect the diversity of Atlanta.

To recruit Atlanta users, we posted a screener survey on Nextdoor, Reddit, and Facebook, as these are sites where there is prior evidence of users engaging with local safety-related information [70, 90, 109]. There were 139 individuals who completed the initial screening survey. We followed up with 67 individuals and invited them for interviews. Twelve of these individuals completed the interviews. All but one of the participants we interviewed found our post on Nextdoor. The majority of people in this sample were between 35 and 44 years old, female, and White. In our second round of recruitment, we aimed to interview more people of color and posted our screening survey on subreddits and Facebook groups for Black colleges in Atlanta. We also posted on two different Nextdoor groups in predominantly Black neighborhoods. There were 72 individuals who completed the recruitment screening survey, 24 of whom self-identified as Black residents of Atlanta. We invited nine of these individuals for interviews, and conducted interviews with the three who accepted.

Participants noted that they had used Citizen between 5 weeks and 2 years, with a rough average of 9.5 months (some participants did not give exact answers). Participants spent between five minutes to 12 hours per week on the app, with a rough average of approximately 87.5 minutes per week (some participants did not give exact answers). Table 1 lists the demographics of all 15 participants. Our participant sample includes the following: 53% of our participants identified as female (n = 8) and 40% identified as male (n = 6). One participant declined to specify their gender. 46.6% identified as White, 6.6% identified as Hispanic or Latino/a, 13.3% as Asian or Pacific Islander, 20% as Black, and 6.6% identified as White and Native American. Additionally, 1 participant declined to specify their race. Despite our targeted recruitment strategy, Black people were underrepresented in our sample. This may be for a number of reasons. Our research team could not find data about the racial makeup of Citizen users to determine whether our participants reflect the broader population of Citizen users in Atlanta, but prior work suggests that Black people are less likely to use social media to find out about local crime activities [56]. Additionally, Black communities in Atlanta have been exploited by researchers and have high levels of distrust which has affected recruitment of this

population in the past [66]. In both the screening survey as well as follow-up emails to schedule the interview, we explained that all interviews would be recorded on Zoom, which may have biased our sample towards those participants who are more trusting of researchers or feel more lax with privacy. There is an opportunity for future research to focus specifically on Black population regarding their usage of the app.

During interviews, we asked participants to describe: (1) the features they used, (2) their motivation for use, (3) how often they used each feature, and (4) their experience with that feature, including the way it may have shaped their behaviors and beliefs. The interviews ranged from 21 minutes to 57 minutes, with the average interview length being 42.31 minutes (sd = 11.04). Each participant was compensated with a $30 e-gift card.

3.2.2 Deceptive Design Pattern Identification. Three members of the research team conducted an interface analysis to identify deceptive design patterns employed by the Citizen app. Adapting a methodology used by Di Geronimo et al. [37] and Gunawan et al. [51], we recorded our interactions with Citizen by following six predefined user scenarios. An iPhone X, an iPhone 13 mini, and a Pixel 4a were used to record and interact with Citizen version 0.1100.0. Researchers recorded the scenarios in their city of residence, which included Atlanta as well as Seattle. Recording incidents in our city of residence was not only practical, but also enabled us to contextualize the incidents we viewed on the app. The six user scenarios were selected to capture the diversity of ways that users can interact with the app, which we learned from user interviews as well as the first author's use of the app for research purposes over the course of one year.

User Scenarios:

(1) **Download and Setup:** Download the application and allow alerts. Share your location data and enter your home address when prompted by the application. Share your contacts, and add 1-2 members of the research team to your Safety Network. Navigate and explore all five tabs at the bottom of the screen. Share, follow, and comment on one incident.
(2) **Incident Alert:** The first time you receive an alert, tap on the alert and explore the landing page. This alert may be about a contact who is added to your Safety Network.
(3) **Random Check:** Explore the Home tab, the Safety Network tab, and the Notifications tab. Customize the settings to your preference.
(4) **Broadcast Live Incident:** Navigate to the Broadcast tab. Give the application permission to use the microphone and camera and start recording a live incident happening in the area (e.g. police cars or helicopters overhead). Submit the incident for moderators to review and stop recording.
(5) **Premium Use:** Upgrade to the Citizen Protect feature and sign up for the free 30-day trial.
(6) **Delete and End Use:** Turn off notifications and delete friends from your Safety Network. Cancel the Citizen Protect subscription. Delete account and remove the application from the phone.

After recording our interactions with the app, we had a total of 18 videos with an average length of 3.35 minutes. We used an inductive approach [33] to identify deceptive design patterns since prior work has not yet examined deceptive design patterns in safety technologies. Using Mathur et al.'s definition of deceptive design patterns [75] and a coding methodology adapted from Radesky et al. [87], three researchers independently watched the videos and identified instances of monetization and reinforcement techniques which we believed modified the underlying choice architecture for us as users. After removing duplicates, we had a total of 34 usage experiences where we believed the design modified the user's choice architecture. It is important to note that we did not consider designer intent during this review– as Di Geronimo and colleagues note, "understanding designers' intentions and ethical decisions is subjective and may lead to imprecision" [37, p.4]. Instead, we chose to assess what was presented in the user interface and whether or not those designs modified the choice architecture for users [75].

3.3 Data Analysis

Our data analysis process occurred in three stages: 1) analysis of the interview data; 2) analysis of the data from the interface review; and 3) integration of the two datasets.

To identify themes in the first twelve interview transcripts, four members of the research team, including the first author, independently coded the transcripts using Delve Tool [35]. The research team met for two weeks to develop the codebook – all disagreements were resolved through discussion. The first author grouped these codes into larger themes. Over the course of five weeks, our research team met weekly to discuss, refine, and iterate on the codes as well as the emerging themes. After collecting our second round of interview data, our team members coded the transcripts using the existing codebook. During this second round, we generated one new code which led us to re-code older transcripts with this new code in mind. At the end of data analysis, we had 36 codes which were grouped into six overarching themes.

To identify deceptive design patterns, the three members of the team who collected and identified the usage experiences organized these usage experiences using affinity diagramming [52] in Miro Board [77]. Affinity diagramming is an inductive approach that allows users to iteratively group data by theme. This process helped us identify six underlying deceptive design patterns which motivated the usage experiences. We renamed these six patterns using existing nomenclature by consulting deceptive design pattern taxonomies from attention capture [79], e-commerce [74], and privacy [20] domains. The final set of six deceptive design patterns and examples of corresponding usage experiences are presented in Table 2.

The first author integrated the two datasets by iteratively matching on 1) feature and 2) concept. For example, interview data that discussed the Safety Network was integrated with data from the interface analysis related to the Safety Network, and interview data that discussed the concept of community was integrated with data from the interface analysis that was related to the community. After this matching process, two other members of the research team provided feedback on the integrated data. Collection and analysis of the two datasets occurred independently, and thus, not all deceptive design patterns were reflected in the user interviews, and not all user experiences were influenced by deceptive design patterns. We present our integrated data in the Results Section, sharing the

deceptive design patterns identified by researchers as well as how those features did and did not influence the user experience.

4 RESULTS

The Citizen interface creates an inflated sense of danger while simultaneously positioning itself as a solution to that danger. We describe the interface components that create this effect and report on users' experiences with these features.

4.1 Manufacturing Anxiety

The Citizen interface presents a stream of incidents that systematically include categories of events that do not pose a risk to the user. Participants consistently told us that they valued using Citizen but felt an increased sense of fear as a result of their engagement with the app. We document how the notification stream, lack of contextual detail, and lack of community contributed to their increased sense that danger lurked around every corner.

4.1.1 Interface Analysis: Indiscriminately Raising the Salience and Visibility of Safety Incidents. In reviewing the interface, we encountered five types of incidents that were shared with users but did not present a threat to their safety. First, the app notified users about incidents that were not proximate. For example, in one instance, the notification feed displayed an incident about a missing child from a neighboring state (see Figure 2e), and in another, it showed mass shootings from another part of the country. These incidents informed users about alarming incidents that were too far away to affect their personal safety but were presented alongside incidents that occurred nearby, expanding the set of alarming events that were shared with users.

Second, we encountered incidents that were not a threat to public safety and represented minimal or no risk to those who were not directly involved. For example, one incident alerted users of an "Occupied Stuck Elevator" (see Figure 2d). Third, we found that incidents persisted on the feed long after they were over. For example, as shown in Figure 2f, users were shown information about a "Small Brush Fire" that had been extinguished nine hours prior. Videos shared on the Live Broadcast Tab appeared to persist for 24 hours, even if the incident had been resolved.

Fourth, the app encouraged users to add friends to their Safety Network (see Figure 3c), and upon doing so, people began receiving intermittent alerts about incidents that the app framed as relevant to their friends' safety. For example, the first author received notifications that a friend was 0.5 miles away from a reported structure fire and, later, that the same friend was 1.1 miles away from a man reported to be armed with a gun and involved in a dispute (see Figure 3d). In a dense metropolitan city where nearly half of all adults live in a home with a gun [57], this may always be the case, but the alerts signaled to the user that there was reason to be concerned for the safety of a loved one, regardless of whether or not that loved one was actually in danger.

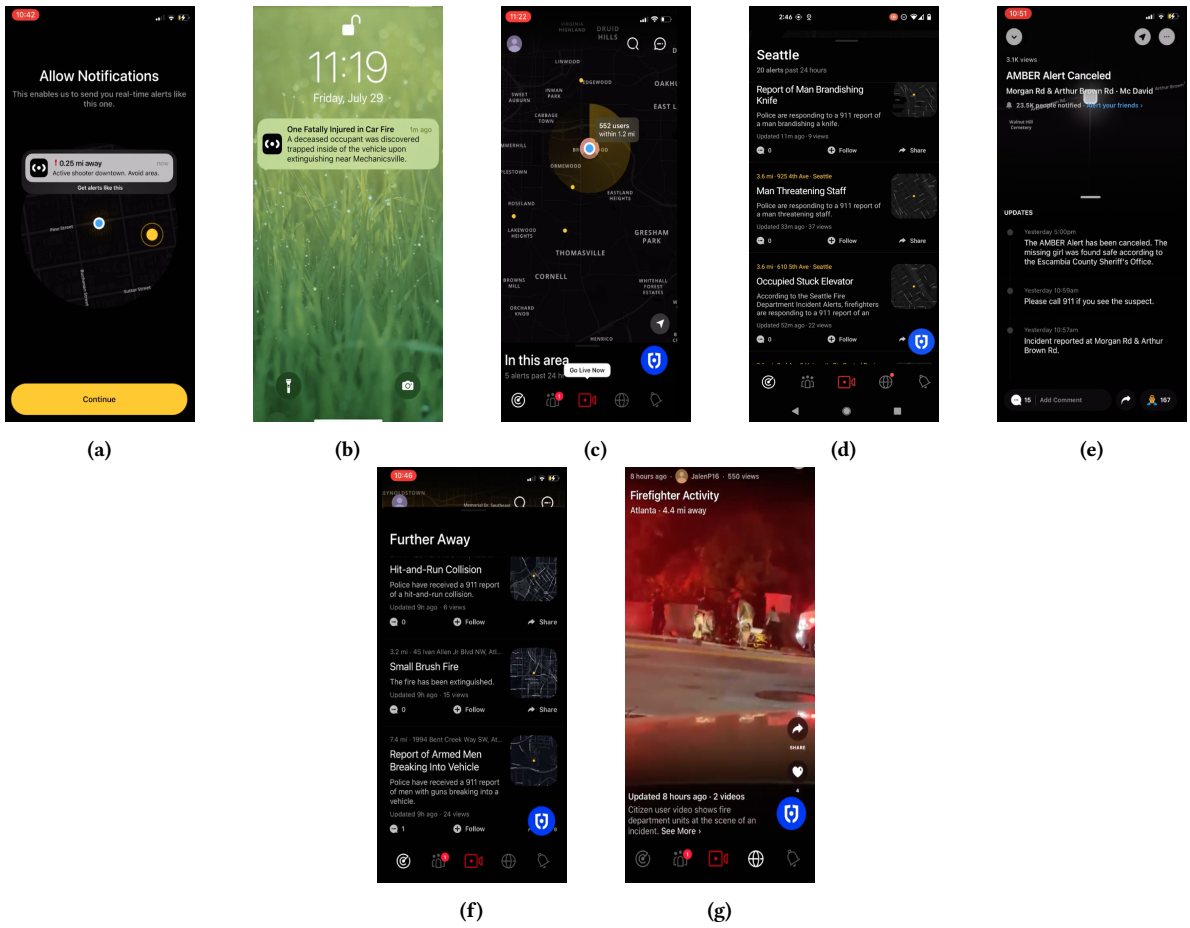
Finally, we encountered incidents that did not provide enough information to determine whether or not the incident presented a safety threat. For example, one incident reported a "Man Threatening Staff" without additional context, leaving the user unsure of how, if at all, the incident related to broader public safety concerns (see Figure 2d). Thus, the collective set of incidents documented events that might be reported as local news stories with few presenting a plausible threat to the user's safety. However, Citizen did not encourage users to consume content as local news; the app encouraged users to stay vigilant and maintain real-time awareness of safety risks like "active shooters" by enabling alerts (see Figure 2a). Citizen required users to enable alerts in order to view their Notification Feed, manufacturing an artificial dependency, what Mathur et al. call a *forced action* deceptive design pattern [74] (all deceptive design patterns are documented in Table 2).

4.1.2 User Experience: Constant Notifications Manufacture Anxiety. All participants reported that Citizen increased their awareness of safety-related incidents in Atlanta. P10 described the app as an "*electronic bubble of information*" that heightens his awareness of his surroundings no matter where he goes. Citizen left participants feeling shocked at how many criminal incidents occur in the city, commenting on the number of car thefts (P8), fires (P11), and instances of gun violence (P10). They expressed their dismay over the prevalence of danger saying things like, *"there's so much crime and you just don't expect that"* (P8), and they explained that this awareness developed through their use of Citizen, which had surfaced a backdrop of crime they had not previously realized existed. For example, P1 told us, *"there's a level of ignorance is bliss where, if you don't know anything going on, you know everything seems safe and happy and then, when you add Citizen suddenly you're aware that there's danger around."*

Participants in our study viewed information from Citizen as reliable because of its unfiltered nature. They trusted the reports because they came from police radio (P8, P12) and perceived incidents to be devoid of the extra commentary (P1), *"sensation"* (P5), and political slant (P7) that they associated with local news and social media posts (P10, P8, P11, P7, P1, P5).

However, the affordance that participants found most valuable was Citizen's ability to provide hyper-local, real-time information. Participants P3, P10, and P4 all shared that they were alerted about incidents that they could see happening outside their house or gunshots that they could hear in their neighborhood, incidents they perceived as relevant to their safety but too minor to be reported on the news. P10 shared that these alerts helped him *"know what to do"* and which places to avoid at what time, and P13 liked that he can find out about crime *"immediately"*. P5 put it succinctly, *"I just want to know, like locally, just straight up what's going on near me."* As these examples illustrate, for our participants, the core use case for Citizen is to cultivate a real-time awareness of nearby events that might affect their safety.

Although participants appreciated the increased awareness that came with using Citizen, they also said that the frequency of alerts was "*stressful*" (P9) and "*anxiety-inducing*" (P6). This is consistent with what users have shared on product reviews of the app [59]. Participants received five to fifteen alerts per day, with the influx becoming "*really crazy*" at night (P8). The incidents that participants felt were the least helpful were ones that were "*far, far away*" (P3) or inconsequential to their personal safety. P11, for example, guessed that maybe "*one out of 20 [incident notifications] is actually useful*" because "*unless you're within half a mile or a quarter-mile away from me, I really don't care.*" P3 and P12 felt similarly, voicing that it was "*annoying*" (P3) to receive so many "*random notifications about*



(a) (b) (c) (d) (e)

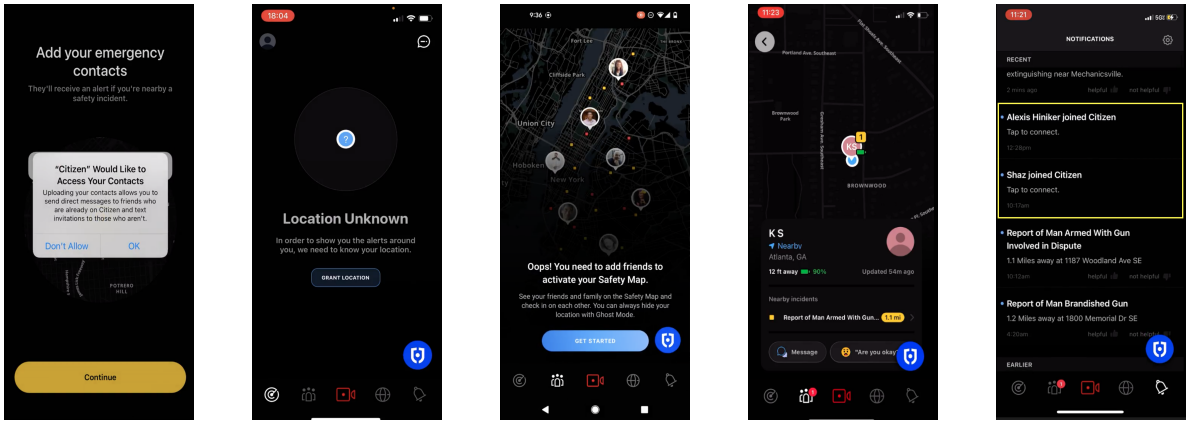
(f) (g)

Figure 2: Users are encouraged to allow incident alerts from the Citizen App (a). Upon receiving an alert (b), users can tap it to view the story, view other nearby alerts on a map (c), or view a list of recent notifications (d). While our research team mostly received local notifications (e.g., within 5 miles), (e) shows a notification about an Amber Alert of a missing child from a different state. (f) displays a list of incidents that are "Further Away" and (g) shows a video on the Live Broadcast tab of a fire that had been resolved hours ago.

things that are not happening within my vicinity" (P12). Participants reported that they often received notifications about "*fires*" (P11) and "*helicopters*" (P9) that they did not care about, and P6 shared that Citizen alerts her about "*a whole bunch of fluff, if you will, you know unnecessary calls to the police.*" Participants expressed frustration with excessive alerts that depict "*all this crime, but it's actually not, and then it makes it not as useful, like the boy who cried wolf*" (P9).

For some participants, the excessive notifications manufactured what they perceived to be an unnecessary sense of fear. For example, P2 explained that "*there's always a little action right around me because I'm by Edgewood, and there's kind of a lot of crap going on in Edgewood, so [the constant stream of notifications] just has me super paranoid.*" Other participants shared that they expected crime in a big city but seeing so much of it was *"scary"* (P4), *"anxiety-inducing"* (P6), and *"not for the faint of heart"* (P8). P6 described this phenomenon by explaining that, because of Citizen, she hears about "*every little teeny tiny thing, whether it's true or not... instead of like hearing the things that actually matter, I see all of these different things that are probably not a concern. But then it's like it's overwhelming to see like, 'wow, 15 different things have happened within a mile from me.'*" P12 agreed, describing the app as *"alarmist"*.

Participants suggested ways that the app might scale back irrelevant notifications and prioritize relevant ones, and as a result, inspire fear only when warranted. For example, P5 reflected on the difference between violent and nonviolent crimes with respect to his safety, and P9 explained that Citizen needed to be more discerning about the "*difference in severity*" between incidents. She wished there were "*more ways to break down when you would get notifications and about what types... like I don't want a notification about a traffic accident but, like, I would like to know if there's a shooting right across the street, or if there was a break in near my complex within you know, a mile or two.*" P11 suggested that Citizen implement a "*geofence*" so that he would only be alerted about notifications that were proximate. Although users did not seem to be aware of the *forced action* deceptive design pattern, the requirement to enable



(a) (b) (c) (d) (e)

Figure 3: During the onboarding process, users are invited to add optional emergency contacts. To do so, they are required to grant Citizen access to their phone's contacts (a). Users are required to share their location in order to access the app's incident map (b) and required to add friends from their contacts to use the Safety Network (c). Once they've added friends to the Safety Network, they can view their friends' current location and battery life on the Safety Network tab (d). When one of their contacts joins Citizen, the user receives a notification prompting them to connect with their contact (e).

alerts in order to view the Notification Feed disempowered users from choosing how and when they would like to view incident information.

4.1.3 User Experience: Lack of Context and Quality-Control Inspires Unwarranted Concern. Participants shared that the lack of contextual information made it difficult to discern whether an incident was cause for concern and wanted Citizen to surface details that would enable them to make this judgment more readily. P1, for example, said,

> "It's very important to be able to separate…what's real, what's a threat, and what's not, because at the end of the day, if you get into a fight with your boyfriend inside your house and you call the police, I'm very sad for you and I hope that you're okay, but, I don't need to see an alert on my app that there is a report of like, you know, 'brawl in the street,' and like, 'someone with a knife chasing a woman' because then I get worried…and so I think that there's a way of making it objective versus just the over-inundation of information that then causes you to not trust it or not wanting to know."

As P1's quote illustrates, participants perceived lack of context made it difficult to differentiate between private incidents and threats to public safety, contributing to unnecessary fear. In other examples, P11 described an incident where the Atlanta Police Department was conducting a drill, but Citizen incorrectly transcribed it as a "*full-on open assault, like, shooting between two different parties*" which led to alarm throughout the neighborhood (as witnessed by P11 in the comments). These kinds of incidents prompted requests for *"quality-control"* (P12) and *"a little fact-checking"* (P9).

Participants also speculated that this lack of quality control, fact-checking, and context led to consuming culturally and racially biased information. P9 reflected on this concern saying, "*The issues of, like, you know somebody like looking into a car, like, I question is like—was it a black person looking into their own car? [Or] was it actually somebody, like, checking car handles and trying to break into cars?*" Similarly, P11 knows "*a lot of 'Karens'* [1] *in the neighborhood*" who are quick to call 911, thereby inflating the Notification Feed with biased incidents that may nevertheless inspire concern.

4.1.4 User Experience: Lack of Community Limits Users' Resilience to Fear. Participants in the study infrequently interacted with other users on Citizen. Twelve participants shared that they had never posted on Citizen. P1 talked about how her contributions to the app consisted of once adding a "*sad emoji*" reaction while P11 described Citizen as a platform where people did not "*make lifelong friendships.*" Participants cited the following deterrents for connecting with other users: personal preference in using the app for quick news alerts (P1), online anonymity (P4), and high amounts of negative content which made it a difficult place to "*hang out*" (P1). Participants said they encountered more community-building activities on other platforms, such as Nextdoor and Facebook. P1 talked about how she turns to Nextdoor and Facebook for "*personal color commentary*" to augment the reports she sees on Citizen. She explained that this commentary enables her to have "*a more complete picture*" of what was happening in her neighborhood, and made it "*a lot easier to live with that danger that you [she] know[s] about from Citizen.*" Reflecting on the impact of the "*personal color commentary*," she shared:

> "It does make it less scary…when you add [start using] Citizen, suddenly you're aware that there's danger around you but you don't know exactly who or how or why that danger exists, you just know that

[1]Karen refers to the 2020 "Karen" meme caricaturing white women who typically overreact and escalate situations including making threats to involve the police or abusing grocery store workers for mask-wearing policies. The meme is often associated with white supremacy [81].

Table 2: Deceptive Design Patterns Employed by Citizen

Deceptive Design Pattern	Definition	Example Within Citizen	Influence on User Experience	Domain
Social Investment	The use of social metrics to reward users for their engagement and incentivize continued use	Sharing the number of nearby users who would presumably view, react, and comment to user-uploaded videos	N/A	Attention Capture [79]
Obstruction	Making a certain action harder than it needs to be in order to dissuade users from taking that action	A hidden "Skip" button needed to avoid premium upgrade	N/A	E-commerce [74]
Misdirection	The use of language, visuals, and emotions to guide users toward or away from making a particular choice	A floating button to upgrade to Citizen Protect overlaid on safety notifications and videos	N/A	E-commerce [74]
Forced Action	Requiring users to take certain tangential actions to complete their tasks	A requirement to enable alerts in order to view the Notification Feed	Users enabled alerts and received information that was not always relevant to their safety concerns	Privacy [74]
Publish	Sharing personal data publicly	An alert that notifies users that a contact has joined Citizen without informing that contact	Users added contacts to their Safety Network and received alerts about contacts that were not always relevant to their safety concerns	Privacy [20]
Obscure	Making it challenging for users to learn how their personal data is collected, stored, and/or processed	The lack of transparency about what personal data is collected and how it is stored	Users did not trust Citizen and felt reluctant to share information with the application	Privacy [20]

Note: Citizen employs known deceptive design patterns from attention capture, privacy, and e-commerce domains.

.

> it is, and then with Nextdoor you get a little more understanding of why this person is waving a gun on the corner and that if you drive through they're not going to shoot out your window, like, they're really pissed at their ex-husband."

As this story illustrates, participants saw the interpersonal communication that occurs on other platforms as humanizing and potentially mitigating fear of crime. One participant shared that the lack of online community on Citizen left few chances for users to "*make sense of what are the motivations and the kinds of things that may be incentivizing that kind of behavior*" leading users on Citizen to be more "*apathetic*" than "*empathetic*" in their comments (P5). While one participant said that he liked that the comments were uncensored (P11), others said they found the comments "*gross*" (P4, P5), "*unkind*" (P4), "*violent*" (P5), and "*racist*" (P11).

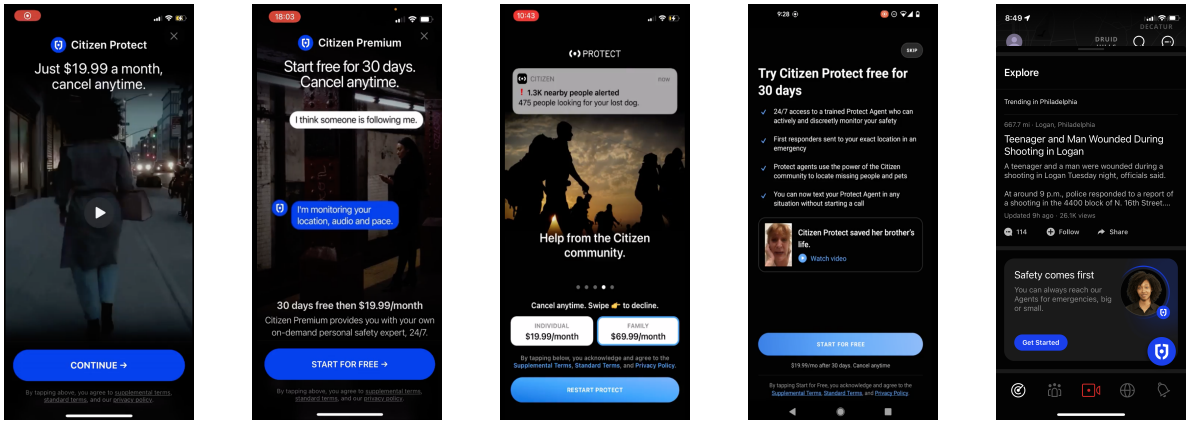
4.2 Offering a Solution to Users' Heightened Safety Needs

Despite the frustration and anxiety that users reported, they also felt it was important to keep using the app to better manage their own safety (P2, P9, P1, P14). The users we interviewed were not unaware of the negative impacts of using Citizen, but felt beholden to the application. Since downloading Citizen, P14 described having a constant urge to know "*what's really going on*" including checking whether a place he is in is "*secure*." P2 shared that she felt "*beholden to these sound alerts that instill panic. It's like Pavlov's dog: you hear the bell and you have a reaction; it's visceral…I feel like a slave to it but it's the only way I'm going to be able to control my safety as much as I can.*" Others agreed—P9 voiced that she has gone back and forth on whether or not to delete the app because it induces anxiety, but decided not to get rid of it because it provided her with valuable information.

Thus, we found that Citizen became both a source of and a solution to anxiety for our participants. Here, we examine the interface features that position Citizen as a solution and the steps—both with and without the app—users took to manage their safety.

4.2.1 Interface Analysis: Encouraging the Use of Lucrative Features Which Promise Protection. Citizen offers users three features for their protection, their loved ones' protection, and their community's protection: Citizen Protect, Safety Network, and Live Broadcast. These three features are also profitable, helping the company gain users' money, data, and attention.

Citizen Protect is Citizen's premium feature which was launched in 2021. The feature offers users the option to contact Citizen employees, known as Protect Agents, who can monitor the user's surroundings, contact first responders when situations escalate, alert users' emergency contacts, and create new incidents on behalf of a user to alert nearby users of the app. Citizen Protect is promoted as a tool that brings people together to watch out for each other



(a) (b) (c) (d) (e)

Figure 4: The Citizen Protect feature is first advertised to users during the onboarding process (a) (b). In-app advertisements highlight the benefits of using Protect and spotlight "success stories" such as finding missing pets (c) or people (d). There is also a floating blue button to sign up for Citizen Protect that is constantly visible in the lower right corner of the screen (e).

[27]. In-app advertisements give the example of a Protect Agent creating an incident to alert nearby users about a missing pet and the nearby users responding en masse (see Figure 4c). Researchers found this vision of mobilizing users reminiscent of Citizen's prior avatar as the Vigilante app. Although the app is free, we found that Citizen aggressively advertises its premium features with the use of deceptive design patterns. For example, the Citizen Protect feature is advertised twice to new users during the sign-up process. In the latter instance, researchers noted the hidden "Skip" button which made it particularly challenging to bypass the advertisement, an example of a deceptive design pattern called *obstruction* [74]. The most egregious deceptive design pattern, however, is a floating button to sign up for Citizen Protect which is overlaid on each screen, constantly visible as users scroll through videos and notifications, many of which do not present any threat to users' safety but heighten fear nonetheless (see Figure 4e). We saw this as an example of a *misdirection* deceptive design pattern, a button which supports Citizen in translating heightened awareness and anxiety about safety into purchases. Users can purchase an individual or a family plan.

Citizen also encourages users to monitor their friends and family's safety by adding contacts to their Safety Network. To take advantage of this feature, Citizen requires users to share their entire contact list with the app (see Figure 3a). There is no option to add contacts individually, an example of a *forced action* deceptive design pattern [74] because it creates a false dependency. If users choose to share their contacts, Citizen will alert all contacts who are existing Citizen users that their friend has joined the app without informing the user. This alert encourages contacts to add the user to their Safety Network and share location data with the user (see Figure 3e). We saw this as an example of *publish*, a privacy deceptive design pattern, [20] where information about an individual is shared without their consent or knowledge. This deceptive design pattern has the potential to exponentially increase new users for Citizen. Researchers also discovered that the app collected data about the user without their knowledge, including data about the user's heart rate and about their mobile device's battery life. Battery life information was shared with friends on the Safety Network without consent. These are examples of privacy deceptive design patterns which *obscure* what data is being collected and how [20].

The app describes Live Broadcast as a feature that allows users to create and share videos in order to "spread awareness of safety incidents with your community in real-time." Citizen nudges users with verbal cues and displays the number of nearby users (who would presumably see the live video) (see Figure 1c, Figure 2c). We see this as an example of a *social investment* deceptive design pattern because it encourages the use of the app through social metrics such as the potential number of reactions, comments, and views to user-uploaded videos [79]. Researchers also documented one instance where users were prompted with the notification: "Go Live. 600 feet away. Hit-and-Run Collision. Tap to update your community" (see Figure 1e). The research team found this notification particularly challenging to reconcile with the app's mission to support user safety [27]. User-generated broadcasts were used to capture and engage users' attention. For example, one researcher received an alert that there was a "live video at the scene", to encourage viewing a video of an overturned car after a collision. Each video was also overlaid with users' comments, reactions, and a pulsating share button to encourage users to share the video via text or social media.

4.2.2 User Experience: A Heightened Need for Safety Requires Action. Sensitized to the risks around them, users engaged Citizen's features for protection in two ways and responded individually, taking matters into their own hands, in many ways.

While we did not speak to any participants who had used Citizen Protect or Live Broadcast and could not evaluate the influence of the *obstruction*, *misdirection*, or *social investment* deceptive design patterns, we did speak to four participants who added friends to their Safety Networks (P1, P3, P4, P6). P6 mentioned that he has a very diverse group of friends, and given the racially-charged political climate, he appreciated the ability to make sure they were safe.

P3 similarly appreciated being able to track her family members' locations. P1 downloaded the Citizen app when her friend invited her to join her Safety Network due to the *publish* deceptive design pattern. While P1 valued the information she received from the app, she decided to turn on "Ghost Mode" because alerts about P1's nearby incidents were causing her friend undue stress and anxiety.

Taking advantage of the information on Citizen, we observed how some participants began engaging in detective work. A Citizen post helped P14, an undergraduate student, create awareness about his missing friend. Other students on his campus also used the app, and P14 found that the comment section provided useful and comforting information when his friend went missing. Some participants viewed incidents on Citizen and cross-referenced that information on other platforms to get more context (P6, P4, P1, P9). P9, for example, was able to collect more information about a neighbor's missing car using Citizen and Facebook, while P4 was able to locate a Nextdoor neighbor's missing mail by cross-referencing information from Citizen.

Others did not feel as comfortable relying on Citizen because they worried about sharing location data with the app (P9, P12, P15, P11). P11 changed his settings so that he was only sharing his location when he was using the app because he assumed Citizen had to make money, and they must be doing something with his data that he was unaware of. P12 lives in an apartment complex where she knows there is gang activity. However, she admitted that she no longer feels comfortable calling 911 because she worries identifiable information might be leaked onto Citizen. She said, "*I can't believe I question now calling 911..it made me think to have like who has access to 911 recordings now?*" Although users did not seem to be aware of specific deceptive design patterns, the lack of transparency about Citizen's privacy policy due to design decisions such as the *obscure* deceptive design pattern disempowered users from taking actions that might protect their safety.

In addition to relying on Citizen, many participants took matters into their own hands and began carrying tasers (P9), guns (P12), knives (P2), mace (P9, P2) and investing in new home security systems (P9, P12, P7). Others began avoiding certain sub-populations perceived as dangerous. A small group of participants shared that their use of the app led to an increased fear of individuals who are homeless (P1), mentally ill (P2), Black teenagers (P2), and "*Black men*" (P4). P12 felt that she sees so many crime-related incidents with such little context that her mind can't help but draw conclusions about who is committing these crimes. P1 reflected that:

> "Before I downloaded Citizen when I would see homeless people in the park I wouldn't think anything of it, you know they're there sleeping, this is a soft relatively private place for you to lay your head tonight, and I would go on my way. Since downloading Citizen, I will leave a little more space, and I will look in those bushes a little more like, 'is there, someone that could potentially be right there waiting to pounce?'"

For P11, Citizen brought to light the city's "*vagrancy problem*" and the sense that more police activity and local leadership is needed.

Almost every participant began avoiding certain areas of the city that they perceived as dangerous. Participants mentioned changing the routes they drove (P8), the routes they walked at night (P2, P6, P9, P11, P4), and the businesses they frequented (P9, P11). Based on the incidents that participants viewed on the app, they began to create mental models of "*hot pockets*" (P6) in the city to avoid. P8, for example, said that after seeing the same street names, again and again, she began avoiding those areas. Similarly, P11 described how he used Citizen to figure out if he should "*avoid that section of town*" for the day. Furthermore, these mental models persisted beyond just the usage of the app. P4, for example, no longer attends the Castleberry art walk because she now associates that neighborhood with crime, and P2 said she no longer goes out for walks alone after six pm. For others, the data from the app has influenced long-term decisions like where to buy a house (P7, P8) and whether it makes sense to move to another state altogether (P10, P2). The areas that participants mentioned as "*hot pockets*" of crime include Castleberry Hill, home to one of the highest concentrations of Black-owned land and businesses in the country, and Mechanicsville, where the vibrant and predominantly Black community of the 1950s has since diminished largely due to misguided urban renewal [1, 5].

5 DISCUSSION

5.1 The Power of Deceptive Infrastructure

In 2021, HCI researchers Westin and Chiasson introduced the concept of *dark digital infrastructure* [108]. They observed that examining individual features in isolation neglects the ways in which these features interact with each other and with larger social and psychological factors [89, 108]. A narrow focus on individual features limits researchers' ability to fully understand the impacts of these designs. To account for this oversight, Westin and Chiasson use "dark infrastructure" to refer to the larger sociotechnical machinery—built on deceptive design patterns—that undermines user agency at scale [108]. In our review of the Citizen app, we similarly observed that a feature-level analysis did not capture the full extent to which Citizen's interface can modify users' choice architecture. Here, we consider how the deceptive design patterns we identified in Citizen might intersect with cognitive biases and sociocultural contexts to produce dark infrastructure. We refer to dark infrastructure as "deceptive infrastructure" to avoid conflating the racialized term "dark" with problematic behavior.

5.1.1 Deceptive Design Patterns and Cognitive Biases. In 2014, Facebook notoriously conducted an experiment to understand how users' emotional states are influenced by the emotional valence of the content on their feeds [61]. While this study was highly controversial, it is not the only example of technology manipulating users' emotional states at scale [107, 108]. Our results describe how Citizen modifies users' choice architecture by sharing information that does not present a threat to users' safety, but heightens anxiety and fear nonetheless. While individual deceptive design patterns (like requiring users to enable notifications) may seem relatively innocuous, we found that over time they created high emotional costs for participants in our study who described their experience as "scary," "anxiety-inducing," "stressful," "frustrating," and "paranoia"-inducing.

Attentional bias is a type of cognitive bias where people disproportionately attend to emotionally evocative information due to

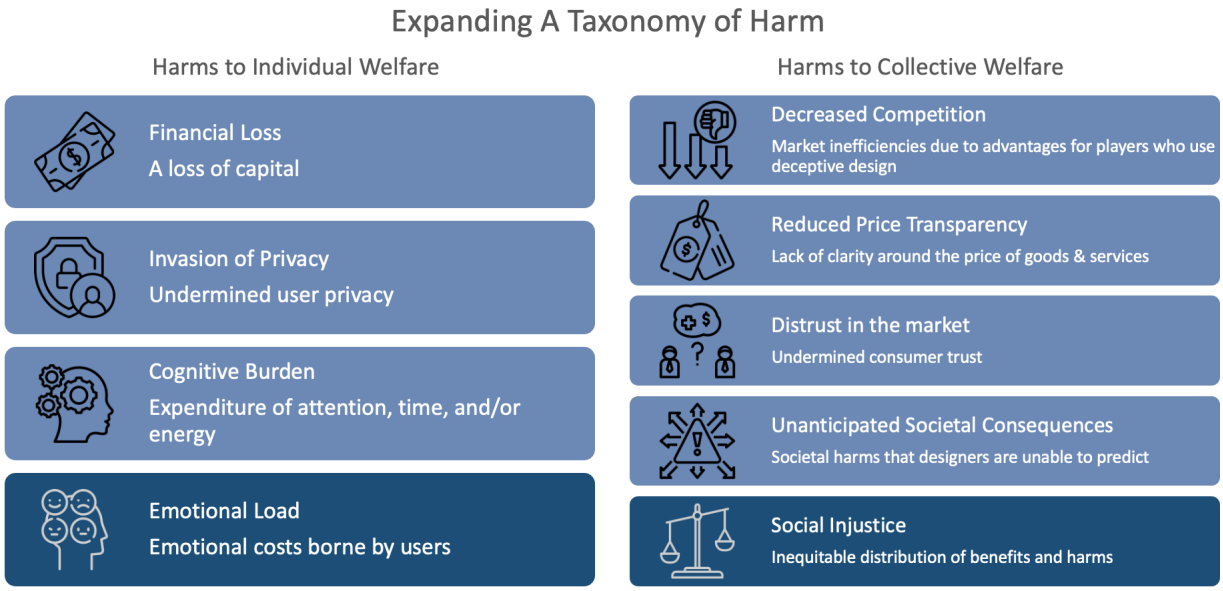


Figure 5: We propose an expansion to the Mathur et al. taxonomy of harm [75]. Light blue items are taken from the existing taxonomy; dark blue items are our proposed additions. All icons are taken from Flaticon [49].

the evolutionary importance of early and fast processing of threat-related information [60]. Given the potential for deceptive design patterns to exploit users' cognitive biases [74, 105], we hypothesize that attentional bias may explain why safety incidents evoked strong emotional responses even when they did not present a threat to user safety. Attentional bias suggests a lowered threshold for modifying users' choice architecture with safety information, and a highly-cited meta-review of attentional bias found that increased exposure to negatively valenced content has a causal, bidirectional, and mutually reinforcing relationship with anxiety [104]. The interactions between individual deceptive design patterns and cognitive biases may thus create a deceptive infrastructure that leaves users vulnerable to manipulation and creates emotional costs that persist even after users log off [96].

5.1.2 Deceptive Design Patterns and Sociocultural Contexts. We identify the potential for deceptive design patterns to interact with the cultural and social contexts within which they exist to systematically reproduce negative stereotypes and reinforce cultural biases. Interviewees reported instances of racism in the comments section of the app and shared that the fear of crime left them feeling increasingly distrustful and suspicious of strangers, particularly Black men and unhoused individuals. This is not surprising given that decades of research has established that the fear of crime is not expressed neutrally, and in the United States, is likely to be directed in ways that reflect existing biases against Black people [40]. The fear of crime is closely associated with a narrative of Black criminality [40, 106], leading to, for example, the policing, profiling, and surveillance of Black, Brown, and low-income populations [12, 45, 63, 70, 82].

Citizen is one of many technologies that interact with their cultural and social contexts in ways that disproportionately impact vulnerable and historically marginalized populations. Technologies used by millions of users to discover new restaurants (Yelp) or buy and sell homes (Zillow) profit off of users' engagement even as that engagement contributes to the reproduction of racial biases and gentrification [29, 114]. Researchers have documented the inconsistent enforcement of online racism on Reddit due to an interest in protecting user engagement on the platform [109] and children from lower socioeconomic backgrounds have been found to play apps with more deceptive designs [87]. These examples all point to the potential for deceptive design patterns to interact with sociocultural contexts to reproduce implicit biases and stereotypes that systematically harm vulnerable and historically marginalized populations.

5.1.3 Expanding the Taxonomy of Harm. In a 2021 meta-review of the literature, Mathur et al. identified individual and collective welfare as overarching normative concerns that underlie the discussion on deceptive design patterns [75]. They offer a taxonomy of harms organized under these two categories with the hope of providing researchers with a common language to explain why deceptive design patterns are of import and concern [75]. Their review of the literature finds that deceptive design patterns have the potential to harm individual welfare through financial loss, invasion of privacy, and cognitive burdens. They also have the potential to harm collective welfare through decreased competition, reduced price transparency, distrust in the market, and unanticipated societal consequences (see Figure 5).

In light of our results, we propose the need to expand this taxonomy to include *emotional load* as harm to individual welfare. Emotional load is defined as the emotional cost borne by users due to a technology's deceptive infrastructure. We see the need for

researchers to begin systematically documenting this harm; leveraging empirically validated measures from psychology to identify and measure complex emotions such as fear can support researchers in this endeavor. As an example, Westin and Chiasson use an empirically validated scale to measure users' "fear of missing out" and the role that deceptive infrastructure plays in producing this fear [108].

Unlike individual welfare which has been a core focus for deceptive design pattern research, collective welfare has received little attention [75]. This is an oversight given the ways that technologies can interact with social and cultural contexts to reproduce harm for whole sub-populations. We propose an expansion of Mathur et al.'s taxonomy to include *social injustice* as harm to collective welfare. Social injustice refers to the inequitable distribution of harms and benefits in society [39]. This is distinct from harm due to *unanticipated societal consequences*, which are harms that designers are unable to predict. Mathur et al. give the example of Cambridge Analytica's use of personal data from Facebook to initiate a disinformation campaign to influence the 2016 U.S. presidential election [75] as an unanticipated societal consequence. In contrast, social injustice can be identified and documented in design using frameworks such as those proposed by Costanza-Chock [30] and Dombrowski et al. [39]. As an example, Corbett engages a social justice framework proposed by Dombrowski et al. to identify the ways that commercially available technologies can reproduce and resist gentrification [29]. By expanding the taxonomy of harm to include social injustice, we hope to draw attention to the ways that deceptive infrastructure can contribute to harm to some populations while benefiting others.

The Federal Trade Commission (FTC) is an independent government agency whose mission is to promote competition and protect consumers from unfair or deceptive practices [102]. It has a long history of investigating and regulating seller behavior which "unreasonably creates or takes advantage of an obstacle to the free exercise of consumer decision-making" [11]. In such cases, the FTC evaluates whether the seller's behavior "causes or is likely to cause substantial injury to consumers which is not reasonably avoidable by consumers themselves and not outweighed by countervailing benefits to consumers or to competition" [2]. Not only does the FTC have the authority to regulate such behavior, but it can also provide remedies for "significant injuries," such as financial losses [86]. In recent years, the FTC has requested feedback on proposals to regulate how companies collect and store user data [3] and how they hide fees [4], for example. By expanding the taxonomy of harm, we hope to raise these issues as critical for researchers to document and for the FTC to start investigating.

5.2 Dismantling the Deceptive Infrastructure of Safety

With a heightened awareness of how deceptive design patterns interact with human biases and sociocultural contexts, designers can better account for the potential harm caused by the technologies they create. In this section, we offer recommendations that demonstrate how an awareness of deceptive infrastructure can be translated into concrete design suggestions. These suggestions would not have been possible if we had focused narrowly on feature-level patterns.

(1) *Empowering Users to Selectively Engage With Safety Information.* Because users disproportionately attend to safety information even if it does not present a threat to their safety [115], it is critical that design empowers users to selectively engage. Participants voiced a need to filter the information they received. They suggested features like a geofence as well as the ability to discriminate between violent and nonviolent incidents. Safety applications could additionally provide users the option to filter for ongoing incidents or ones occurring in public rather than in private spaces. Furthermore, implementing processes to verify reported incidents and sharing that process transparently can help users assess the potential threat an incident poses.
(2) *Contextualizing Danger Over Place and Time.* Presenting an authoritative and singular representation of place that is governed entirely by crime can make it easy for users to feel scared and default to unexamined assumptions about a place and the people who live there [62]. Providing users with feeds that reflect not just crime, but a diversity of events can help users maintain perspective. For example, alongside stories that highlight criminal incidents, platforms could also share community events, highlight instances of collaboration, or celebrate individual members of the community. This type of diversity can support users in developing a more nuanced understanding of place and people. Furthermore, longitudinal data can help contextualize individual incidents. For example, property crime in Atlanta has decreased steadily and dramatically between 2009 and 2021, and violent crime has decreased significantly since 2009, with a slight uptick between 2018 and 2021 that nevertheless remains lower than any year 2017 or earlier [53]. Design which communicates these longitudinal trends can support users in contextualizing safety incidents within a longer time frame.
(3) *Actively Dismantling Cultural Stereotypes.* Decades of research on implicit biases and cultural stereotypes have documented the ways that Blackness is associated with criminality in US culture [34, 40, 106]. Black people are more likely to be characterized by White people as violent and perceived as more likely to engage in criminal activity than White people [101]. As evidenced by Facebook [90, 91], Reddit [109], Nextdoor [63, 70], and WhatsApp [78], safety technologies that do not actively engage with these stereotypes risk reproducing them. Design, however, can play an active role in dismantling cultural stereotypes through the use of evidence-based strategies, such as by promoting counter narratives and embedding opportunities for media literacy training [88]. Prior research by Jessa Dickinson and colleagues, for example, have designed safety technology for street outreach workers to support the dissemination of counter-narratives [38].
(4) *Channeling Fear Productively.* Engaging with safety information is likely to inspire fear [104], but that fear can be

channeled in ways that strengthen the community. Collective efficacy is a measure of a community's level of social cohesion and how effectively that social cohesion can be mobilized towards a common goal [43, 94, 95]. Collective efficacy is a robust predictor of lower rates of violence [95] and increasing collective efficacy is a well-established strategy when designing for community crime prevention [42, 43, 58, 69]. Supporting collective efficacy both online and offline can empower users to channel their fear in productive ways. For example, designers can offer features to organize support for victims after a safety incident or features that encourage users to connect with local nonprofits. These evidence-based strategies both increase collective efficacy and empower individuals to channel fear towards efforts that strengthen a community [43, 93, 100]. Without such channels, users default to individual responses which ultimately create suspicion and distrust [28] and decrease feelings of safety [98].

5.3 Limitations

There are a few limitations of this study. First, 12 of the 15 user interviews were done almost a year prior to the interface analysis of the application. While the main functionalities of the app remained the same, the first author who used Citizen for the duration of the study did note some incongruence. For example, by the time we conducted the interface analysis, Citizen appeared to be advertising Citizen Protect more aggressively and using nudges to encourage users to Live Broadcast. We hypothesize that the reason we were not able to interview users who had broadcasted or used Citizen Protect is that these were not popular features at the time the interviews were conducted, due to their limited advertising. This data would have further illuminated the ways that deceptive design patterns can create purchase pressure. Future work could contribute meaningfully by taking a longitudinal approach to understanding how the influence of deceptive design patterns evolves over time.

A second limitation is the lack of precision in understanding users' emotional states. Participants used words like stress, worry, insecurity, anxiety, fear, and paranoia interchangeably, limiting our ability to specify the exact nature of the user experience. For this reason, we suggest future work draw on methods from psychology to precisely define the influence technologies have on users' emotional states.

Third, our findings are unique to Atlanta users in 2020 and 2021. Users from different cities at different time periods may have very different experiences with the application. Since companies often conduct A/B testing which provides some users with views that differ from views presented to other users, even the participants we spoke to may have had different views of the application. For this reason, we suggest that future investigations of deceptive design patterns using case methods clearly communicate the bounds of the case and refrain from generalizing beyond those bounds.

Fourth, consistent with prior literature on deceptive design patterns [37, 74, 87], the research team conducted an interface analysis of the Citizen app. However, this approach likely limited the number of patterns that we were able to identify since the review was restricted by the experience of three users. Further, because the user interviews were conducted prior to the interface analysis, we may have attended more to features that were discussed by users, including incident alerts and feeds. Future work can account for these limitations by supplementing researcher reviews with users' posts and comments directly from the application. This may be especially useful to identify deceptive design patterns in domains that are understudied.

Finally, as with other interview-based research, our data is self-reported. Participants could have misremembered, selectively shared information, or may have interpreted past experiences and emotions differently than how they were originally experienced. Participants may have been especially hesitant to share the negative influences of Citizen on their emotional states or behavior due to the heightened vulnerability that such responses demand.

6 CONCLUSION

Our goal in this paper was to investigate how deceptive design patterns manifest in safety technologies and how they influence the user experience. We conducted a case study of the Citizen app, a commercially-available location-based crime alert technology. By triangulating interview data with an interface review of the app, we find that feature-level deceptive design patterns interact with sociocultural factors and cognitive biases to create unique harms to both individual and collective welfare. This work contributes to an emerging discussion about deceptive infrastructure. We propose an expansion to Mathur et al.'s existing taxonomy of harm to include *emotional load* and *social injustice* and offer suggestions for designers interested in dismantling the deceptive infrastructure of safety technologies.

ACKNOWLEDGMENTS

We thank the participants for their candidness and their time. We additionally appreciate the anonymous reviewers and members of the Georgia Tech Public Participation Lab (Christopher, Alyssa, Ashley, and Meichen) for their ideas and edits. We finally thank Sunny, Saharsh, Shipra, and Tarun Chordia as well as Kartik Shastri for user testing and unwavering support.

REFERENCES

[1] 2021. BNC Raises The Bar On Juneteenth Coverage To Create Premier TV Destination For Emancipation Day Celebrations. (2021). https://apnews.com/article/juneteenth-lifestyle-3103c12dc772aea12121dfa833bfeb06
[2] 2021. A Brief Overview of the Federal Trade Commission's Investigative, Law Enforcement, and Rulemaking Authority. https://www.ftc.gov/about-ftc/mission/enforcement-authority
[3] 2022. Trade Regulation Rule on Commercial Surveillance and Data Security. https://www.federalregister.gov/documents/2022/08/22/2022-17752/trade-regulation-rule-on-commercial-surveillance-and-data-security
[4] 2022. Unfair or Deceptive Fees Trade Regulation Rule Commission Matter No. R207011. https://www.federalregister.gov/documents/2022/11/08/2022-24326/unfair-or-deceptive-fees-trade-regulation-rule-commission-matter-no-r207011
[5] n.d.. History: Mechanicsville. (n.d.). http://mechanicsvilleatl.org/history/
[6] Census 2021. 2021. QuickFacts Atlanta city, Georgia. (2021). https://www.census.gov/quickfacts/atlantacitygeorgia
[7] David Abrams. 2021. City Crime Stats Crime in Major U.S. Cities. https://citycrimestats.com/covid/
[8] Boone Ashworth. 2021. What is Citizen's criteria for reporting incidents? (Aug. 2021). https://www.wired.com/story/citizen-protect-subscription/
[9] Trevor Bach. 2020. The 10 U.S. Cities With the Largest Income Inequality Gaps. (2020). https://www.usnews.com/news/cities/articles/2020-09-21/us-cities-with-the-biggest-income-inequality-gaps

[10] Trevor Bach. 2020. The 10 U.S. Cities With the Largest Income Inequality Gaps. (2020). https://www.usnews.com/news/cities/articles/2020-09-21/us-cities-with-the-biggest-income-inequality-gaps

[11] J Howard Beales III. 2003. The Federal Trade Commission's use of unfairness authority: Its rise, fall, and resurrection. *Journal of Public Policy & Marketing* 22, 2 (2003), 192–200.

[12] Ruha Benjamin. 2019. Race after technology: Abolitionist tools for the new jim code. *Social forces* (2019).

[13] Steven Bertoni. 2019. Murder! Muggings! Mayhem! How An Ex-Hacker Is Trying To Use Raw 911 Data To Turn Citizen Into The Next Billion-Dollar App. *Forbes* (july 2019). https://www.forbes.com/sites/stevenbertoni/2019/07/15/murder-muggings-mayhem-how-an-ex-hacker-is-trying-to-use-raw-911-data-to-turn-citizen-into-the-next-billion-dollar-app/?sh=5a3e2dd21f8a

[14] Jan Blom, Divya Viswanathan, Mirjana Spasojevic, Janet Go, Karthik Acharya, and Robert Ahonius. 2010. Fear and the city: role of mobile services in harnessing safety and security in urban use contexts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1841–1850.

[15] Mark A Blythe, Peter C Wright, and Andrew F Monk. 2004. Little brother: could and should wearable computing technologies be applied to reducing older people's fear of crime? *Personal and Ubiquitous Computing* 8, 6 (2004), 402–415.

[16] Terah Boyd and Dave Huddleston. 2020. Metro leaders have mixed reaction to public safety app that lets you stream crime right to police. *WSB-TV* (2020). https://www.theguardian.com/world/2017/mar/12/netherlands-will-pay-the-price-for-blocking-turkish-visit-erdogan

[17] Lauren Bridges. 2021. Infrastructural obfuscation: unpacking the carceral logics of the Ring surveillant assemblage. *Information, Communication & Society* 24, 6 (2021), 830–849.

[18] Harry Brignull. 2022. About. https://www.deceptive.design/

[19] AJ Bernheim Brush, Jaeyeon Jung, Ratul Mahajan, and Frank Martinez. 2013. Digital neighborhood watch: Investigating the sharing of camera data amongst neighbors. In *Proceedings of the 2013 conference on Computer supported cooperative work*. 693–700.

[20] Christoph Bösch, Benjamin Erb, Frank Kargl, Henning Kopp, and Stefan Pfattheicher. 2016. Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns. *Proceedings on Privacy Enhancing Technologies* 2016, 4 (Oct. 2016), 237–254. https://doi.org/10.1515/popets-2016-0038

[21] M. Ryan Calo. 2013. Digital Market Manipulation. *SSRN Electronic Journal* (2013). https://doi.org/10.2139/ssrn.2309703

[22] Ryan Calo and Alex Rosenblat. 2017. The Taking Economy: Uber, Information, and Power. *SSRN Electronic Journal* (2017). https://doi.org/10.2139/ssrn.2929643

[23] Citizen. 2021. How do I enable notifications and location? (2021). https://support.citizen.com/hc/en-us/articles/115000606974-How-do-I-enable-notifications-and-location-

[24] Citizen. 2021. What is Citizen's criteria for reporting incidents? (2021). https://support.citizen.com/hc/en-us/articles/115000603373-What-is-Citizen-s-criteria-for-reporting-incidents-

[25] Citizen. 2021. Where is Citizen available? (2021). https://support.citizen.com/hc/en-us/articles/115000273653-Where-is-Citizen-available-

[26] Citizen App Frequently Asked Questions 2021. How does Citizen work? https://support.citizen.com/hc/en-us/articles/115000278894-How-does-Citizen-work-

[27] CitizenAbout 2022. About. https://citizen.com/about

[28] John E Conklin. 1975. *The impact of crime.* Macmillan New York.

[29] Eric Corbett and Yanni Loukissas. 2019. Engaging gentrification as a social justice issue in HCI. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–16.

[30] Sasha Costanza-Chock. 2020. *Design justice: Community-led practices to build the worlds we need.* The MIT Press.

[31] Atlanta Anti-Violence Advisory Council. 2021. *2021 ANTI-VIOLENCE ADVISORY COUNCIL RECOMMENDATIONS REPORT.* Technical Report. City of Atlanta, GA Office of Communications, Atlanta, GA. https://www.atlantaga.gov/home/showdocument?id=51962

[32] Norwegian Consumer Council. 2018. Deceived by design, how tech companies use dark patterns to discourage us from exercising our rights to privacy. *Norwegian Consumer Council Report* (2018).

[33] John W Creswell and Cheryl N Poth. 2016. *Qualitative inquiry and research design: Choosing among five approaches.* Sage publications.

[34] Angela Y Davis. 2011. *Are prisons obsolete?* Seven Stories Press.

[35] Delve Tool 2021. Delve Tool. https://delvetool.com/ software to analyze qualitative data.

[36] Shaila Dewan and B Goodman. 2006. Gentrification changing face of new Atlanta. *New York Times* 11 (2006). https://www.nytimes.com/2006/03/11/us/gentrification-changing-face-of-new-atlanta.html

[37] Linda Di Geronimo, Larissa Braz, Enrico Fregnan, Fabio Palomba, and Alberto Bacchelli. 2020. UI dark patterns and where to find them: a study on mobile applications and user perception. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–14.

[38] Jessa Dickinson, Jalon Arthur, Maddie Shiparski, Angalia Bianca, Alejandra Gonzalez, and Sheena Erete. 2021. Amplifying Community-led Violence Prevention as a Counter to Structural Oppression. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–28.

[39] Lynn Dombrowski, Ellie Harmon, and Sarah Fox. 2016. Social justice-oriented interaction design: Outlining key design strategies and commitments. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. 656–671.

[40] Jennifer L Eberhardt, Phillip Atiba Goff, Valerie J Purdie, and Paul G Davies. 2004. Seeing black: race, crime, and visual processing. *Journal of personality and social psychology* 87, 6 (2004), 876.

[41] Sheena Erete and Jennifer O. Burrell. 2017. Empowered Participation: How Citizens Use Technology in Local Governance. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, Denver Colorado USA, 2307–2319. https://doi.org/10.1145/3025453.3025996

[42] Sheena Lewis Erete. 2013. Protecting the home: exploring the roles of technology and citizen activism from a burglar's perspective. In *Proceedings of the sigchi conference on human factors in computing systems*. 2507–2516.

[43] Sheena L Erete. 2014. Community, group and individual: A framework for designing community technologies. *The Journal of Community Informatics* 10, 1 (2014).

[44] Sheena L Erete. 2015. Engaging around neighborhood issues: How online communication affects offline behavior. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 1590–1601.

[45] Virginia Eubanks. 2018. *Automating inequality: How high-tech tools profile, police, and punish the poor.* St. Martin's Press.

[46] Richard Fausset. 2022. What We Know About the Shooting Death of Ahmaud Arbery. (2022). https://www.nytimes.com/article/ahmaud-arbery-shooting-georgia.html

[47] Bent Flyvbjerg. 2006. Five misunderstandings about case-study research. *Qualitative inquiry* 12, 2 (2006), 219–245.

[48] Partners for Home. 2020. Point-in-Time Count (2020) - Partners For HOME. https://partnersforhome.org/wp-content/uploads/2020/08/2020-PIT-Full-Report_FINAL-1.pdf

[49] noomtah ultimatearm goodware Freepik, Uniconlabs. 2022. Flaticon. https://www.flaticon.com/free-icons/

[50] Saul Greenberg, Sebastian Boring, Jo Vermeulen, and Jakub Dostal. 2014. Dark patterns in proxemic interactions: a critical perspective. In *Proceedings of the 2014 conference on Designing interactive systems*. 523–532.

[51] Johanna Gunawan, Amogh Pradeep, David Choffnes, Woodrow Hartzog, and Christo Wilson. 2021. A Comparative Study of Dark Patterns Across Web and Mobile Modalities. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 1–29. https://doi.org/10.1145/3479521

[52] Gunnar Harboe and Elaine M Huang. 2015. Real-world affinity diagramming practices: Bridging the paper-digital gap. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 95–104.

[53] Moshe Haspel. 2022. *Atlanta Crime Rates in Historical Perspective.* Technical Report. Atlanta Regional Commission, Atlanta, GA. https://33n.atlantaregional.com/friday-factday/atlanta-crime-in-historical-perspective-2009-2021

[54] M. J. Hattingh. 2015. The use of Facebook by a Community Policing Forum to combat crime. In *Proceedings of the 2015 Annual Research Conference on South African Institute of Computer Scientists and Information Technologists - SAICSIT '15*. ACM Press, Stellenbosch, South Africa, 1–10. https://doi.org/10.1145/2815782.2815811

[55] David Ingram and Cyrus Farivar. 2021. Inside citizen: The public safety app pushing surveillance boundaries. https://www.nbcnews.com/tech/tech-news/citizen-public-safety-app-pushing-surveillance-boundaries-rcna1058

[56] Aarti Israni, Sheena Erete, and Che L. Smith. 2017. Snitches, Trolls, and Social Norms: Unpacking Perceptions of Social Media Use for Crime Prevention. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, Portland Oregon USA, 1193–1209. https://doi.org/10.1145/2998181.2998238

[57] Tammy Joyner. 2022. Your guide to Georgia's gun laws. https://atlantaciviccircle.org/2022/05/28/your-guide-to-georgias-gun-laws/

[58] Cristina Kadar, Yiea-Funk Te, Raquel Rosés Brüngger, and Irena Pletikosa Cvijikj. 2016. Digital Neighborhood Watch: To Share or Not to Share?. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, San Jose California USA, 2148–2155. https://doi.org/10.1145/2851581.2892400

[59] Liam Kennedy and Madelaine Coelho. 2022. Security, Suspicion, and Surveillance? There's an App for That. *Surveillance & Society* 20, 2 (2022), 127–141.

[60] Ernst HW Koster, Geert Crombez, Stefaan Van Damme, Bruno Verschuere, and Jan De Houwer. 2004. Does imminent threat capture and hold attention? *Emotion* 4, 3 (2004), 312.

[61] Adam DI Kramer, Jamie E Guillory, and Jeffrey T Hancock. 2014. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences* 111, 24 (2014), 8788–8790.

[62] Laura Kurgan. 2013. *Close up at a distance: Mapping, technology, and politics.* MIT Press.

[63] Rahim Kurwa. 2019. Building the digitally gated community: The case of Nextdoor. *Surveillance & Society* 17, 1/2 (2019), 111–117.

[64] Cherie Lacey and Catherine Caudwell. 2019. Cuteness as a 'dark pattern'in home robots. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 374–381.

[65] Jamiles Lartey. 2018. Nowhere for people to go: Who will survive the gentrification of Atlanta. *The Guardian* 23 (2018).

[66] Christopher A Le Dantec and Sarah Fox. 2015. Strangers at the gate: Gaining access, building rapport, and co-constructing community-based research. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*. 1348–1358.

[67] Chris Lewis. 2014. *Irresistible Apps: Motivational design patterns for apps, games, and web-based communities*. Springer.

[68] Dan A Lewis and Greta Salem. 2017. Community crime prevention: An analysis of a developing strategy. *The Fear of Crime* (2017), 507–523.

[69] Sheena Lewis and Dan A Lewis. 2012. Examining technology that supports community policing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1371–1380.

[70] Maria R. Lowe, Madeline Carrola, Dakota Cortez, and Mary Jalufka. 2021. "I Live Here": How Residents of Color Experience Racialized Surveillance and Diversity Ideology in a Liberal Predominantly White Neighborhood. *Social Currents* (Dec 2021), 23294965211052544. https://doi.org/10.1177/23294965211052545

[71] Maria R. Lowe, Angela Stroud, and Alice Nguyen. 2017. Who Looks Suspicious? Racialized Surveillance in a Predominantly White Neighborhood. *Social Currents* 4, 1 (Feb 2017), 34–50. https://doi.org/10.1177/2329496516651638

[72] Jamie Luguri and Lior Jacob Strahilevitz. 2021. Shining a light on dark patterns. *Journal of Legal Analysis* 13, 1 (2021), 43–109.

[73] Kai Lukoff, Ulrik Lyngs, Himanshu Zade, J Vera Liao, James Choi, Kaiyue Fan, Sean A Munson, and Alexis Hiniker. 2021. How the design of youtube influences user sense of agency. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–17.

[74] Arunesh Mathur, Gunes Acar, Michael J. Friedman, Elena Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. 2019. Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 1–32. https://doi.org/10.1145/3359183 arXiv:1907.07032 [cs].

[75] Arunesh Mathur, Jonathan Mayer, and Mihir Kshirsagar. 2021. What Makes a Dark Pattern... Dark? Design Attributes, Normative Considerations, and Measurement Methods. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–18. https://doi.org/10.1145/3411764.3445610 arXiv:2101.04843 [cs].

[76] Sharan B Merriam and Robin S Grenier. 2019. *Qualitative research in practice: Examples for discussion and analysis*. John Wiley and Sons.

[77] Miro 2022. https://miro.com/ Visual Collaboration Software.

[78] Anouk Mols and Jason Pridmore. 2019. When citizens are "actually doing police work": The blurring of boundaries in WhatsApp neighbourhood crime prevention groups in The Netherlands. *Surveillance & Society* 17, 3/4 (2019), 272–287.

[79] Alberto Monge Roffarello and Luigi De Russis. 2022. Towards Understanding the Dark Patterns That Steal Our Attention. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. ACM, New Orleans LA USA, 1–7. https://doi.org/10.1145/3491101.3519829

[80] Arvind Narayanan, Arunesh Mathur, Marshini Chetty, and Mihir Kshirsagar. 2020. Dark Patterns: Past, Present, and Future: The evolution of tricky user interfaces. *Queue* 18, 2 (2020), 67–92.

[81] Diane Negra and Julia Leyda. 2021. Querying 'Karen': The rise of the angry white woman. *European Journal of Cultural Studies* 24, 1 (2021), 350–357.

[82] Safiya Umoja Noble. 2018. Algorithms of oppression. In *Algorithms of Oppression*. New York University Press.

[83] Midas Nouwens, Ilaria Liccardi, Michael Veale, David Karger, and Lalana Kagal. 2020. Dark patterns after the GDPR: Scraping consent pop-ups and demonstrating their influence. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13.

[84] Richard A. Oppel Jr., Derrick Taylor, and Nicholas Bogel-Burroughs. 2022. What to Know About Breonna Taylor's Death. (2022). https://www.nytimes.com/article/breonna-taylor-police.html

[85] John Perry and Emily Merwin DiRico. 2021. Atlanta crime trends, 2011-present. *The Atlanta Journal-Constitution* (2021). https://www.ajc.com/news/crime/atlanta-crime-trends-2011-present/XIIM6AMGHBHABMBVHCIHM3AOPQ/

[86] Robert Pitofsky. 1976. Beyond Nader: consumer protection and the regulation of advertising. *Harv. L. Rev.* 90 (1976), 661.

[87] Jenny Radesky, Alexis Hiniker, Caroline McLaren, Eliz Akgun, Alexandria Schaller, Heidi M. Weeks, Scott Campbell, and Ashley N. Gearhardt. 2022. Prevalence and Characteristics of Manipulative Design in Mobile Applications Used by Children. *JAMA Network Open* 5, 6 (Jun 2022), e2217641. https://doi.org/10.1001/jamanetworkopen.2022.17641

[88] Srividya Ramasubramanian. 2007. Media-based strategies to reduce racial stereotypes activated by news stories. *Journalism & Mass Communication Quarterly* 84, 2 (2007), 249–264.

[89] Yvonne Rogers, Margot Brereton, Paul Dourish, Jodi Forlizzi, and Patrick Olivier. 2021. The dark side of interaction design. In *Extended abstracts of the 2021 CHI conference on human factors in computing systems*. 1–2.

[90] Niharika Sachdeva and Ponnurangam Kumaraguru. 2015. Deriving requirements for social media based community policing: insights from police. In *Proceedings of the 16th Annual International Conference on Digital Government Research*. ACM, Phoenix Arizona, 316–317. https://doi.org/10.1145/2757401.2757452

[91] Niharika Sachdeva and Ponnurangam Kumaraguru. 2015. Social networks for police and residents in India: exploring online communication for crime prevention. In *Proceedings of the 16th Annual International Conference on Digital Government Research*. ACM, Phoenix Arizona, 256–265. https://doi.org/10.1145/2757401.2757420

[92] Robert J Sampson. 1988. Local friendship ties and community attachment in mass society: A multilevel systemic model. *American sociological review* (1988), 766–779.

[93] Robert J Sampson. 2017. Collective efficacy theory: Lessons learned and directions for future inquiry. *Taking stock* (2017), 149–167.

[94] Robert J Sampson, Jeffrey D Morenoff, and Felton Earls. 1999. Beyond social capital: Spatial dynamics of collective efficacy for children. *American sociological review* (1999), 633–660.

[95] Robert J Sampson, Stephen W Raudenbush, and Felton Earls. 1997. Neighborhoods and violent crime: A multilevel study of collective efficacy. *science* 277, 5328 (1997), 918–924.

[96] Lisette J Schmidt, Artem V Belopolsky, and Jan Theeuwes. 2017. The time course of attentional bias to cues of threat and safety. *Cognition and Emotion* 31, 5 (2017), 845–857.

[97] ND Schüll. 2014. Addiction by Design: Machine Gambling in Las Vegas Princeton.

[98] Mary E Schwab-Stone, Tim S Ayers, Wesley Kasprow, Charlene Voyce, Charles Barone, Timothy Shriver, and Roger P Weissberg. 1995. No safe haven: A study of violence exposure in an urban community. *Journal of the American Academy of Child & Adolescent Psychiatry* 34, 10 (1995), 1343–1352.

[99] Sumit Shah, Fenye Bao, Chang-Tien Lu, and Ing-Ray Chen. 2011. CROWDSAFE: crowd sourcing of crime incidents and safe routing on mobile devices. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '11*. ACM Press, Chicago, Illinois, 521. https://doi.org/10.1145/2093973.2094064

[100] Patrick Sharkey, Gerard Torrats-Espinosa, and Delaram Takyar. 2017. Community and the crime decline: The causal effect of local nonprofits on violent crime. *American Sociological Review* 82, 6 (2017), 1214–1240.

[101] Lee Sigelman and Steven A. Tuch. 1997. Metastereotypes: Blacks' Perceptions of Whites' Stereotypes of Blacks. *The Public Opinion Quarterly* 61, 1 (1997), 87–101. http://www.jstor.org/stable/2749513

[102] the Premerger Notification Office Staff, DPIP Staff, and CTO. 2022. Mission. https://www.ftc.gov/about-ftc/mission

[103] Elliot Tan, Huichuan Xia, Cheng Ji, Ritu Virendra Joshi, and Yun Huang. 2015. Designing a mobile crowdsourcing system for campus safety. *iConference 2015 Proceedings* (2015).

[104] Bram Van Bockstaele, Bruno Verschuere, Helen Tibboel, Jan De Houwer, Geert Crombez, and Ernst Koster. 2014. A review of current evidence for the causal impact of attentional bias on fear and anxiety. *PSYCHOLOGICAL BULLETIN* 140, 33 (2014), 682–721. https://doi.org/10.1037/a0034834

[105] Ari Ezra Waldman. 2020. Cognitive biases, dark patterns, and the 'privacy paradox'. *Current Opinion in Psychology* 31 (Feb. 2020), 105–109. https://doi.org/10.1016/j.copsyc.2019.08.025

[106] Kelly Welch. 2007. Black criminal stereotypes and racial profiling. *Journal of contemporary criminal justice* 23, 3 (2007), 276–288.

[107] Fiona Westin and Sonia Chiasson. 2019. Opt out of privacy or" go home" understanding reluctant privacy behaviours through the FoMO-centric design paradigm. In *Proceedings of the New Security Paradigms Workshop*. 57–67.

[108] Fiona Westin and Sonia Chiasson. 2021. "It's So Difficult to Sever that Connection": The Role of FoMO in Users' Reluctant Privacy Behaviours. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–15. https://doi.org/10.1145/3411764.3445104

[109] Qunfang Wu, Louisa Kayah Williams, Ellen Simpson, and Bryan Semaan. 2022. Conversations About Crime: Re-Enforcing and Fighting Against Platformed Racism on Reddit. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (Mar 2022), 1–38. https://doi.org/10.1145/3512901

[110] Robert K Yin. 2011. *Applications of case study research*. sage.

[111] José P Zagal, Staffan Björk, and Chris Lewis. 2013. Dark patterns in the design of games. In *Foundations of Digital Games 2013*.

[112] Min Zhang, Arosha K Bandara, Blaine Price, Graham Pike, Zoe Walkington, Camilla Elphick, Lara Frumkin, Richard Philpot, Mark Levine, Avelie Stuart, et al. 2020. Designing Technologies for Community Policing. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–9.

[113] zoom 2021. Zoom(Version 5.4.9). https://zoom.us. Video Conferencing Software.

[114] Sharon Zukin, Scarlett Lindeman, and Laurie Hurson. 2017. The omnivore's neighborhood? Online restaurant reviews, race, and gentrification. *Journal of Consumer Culture* 17, 3 (2017), 459–479.

[115] Ariel Zvielli, Amit Bernstein, and Ernst HW Koster. 2015. Temporal dynamics of attentional bias. *Clinical Psychological Science* 3, 5 (2015), 772–788.